

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche

Corso di Laurea Magistrale in  
Scienze Statistiche



**SORVEGLIANZA ONLINE NON PARAMETRICA DI  
VIDEO FRAME PARZIALMENTE OSSERVABILI**

Relatrice: Prof.ssa Giovanna Capizzi  
Dipartimento di Scienze Statistiche

Laureanda: Mariafrancesca Patalano  
Matricola n. 2045237

Anno Accademico 2022/2023



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 La sorveglianza statistica dei dati ad elevata dimensionalità</b>	<b>3</b>
1.1 Introduzione alla sorveglianza statistica . . . . .	3
1.2 La sorveglianza sequenziale dei flussi di immagini . . . . .	7
1.2.1 Sorveglianza e riduzione della dimensionalità . . . . .	8
1.2.2 Sorveglianza in presenza di osservazioni parziali . . . . .	11
<b>2 Sorveglianza sequenziale e strategie di campionamento</b>	<b>19</b>
2.1 Generalità sulle strategie di campionamento . . . . .	19
2.2 <i>Thompson Sampling</i> per il problema del <i>Multi-Armed Bandit</i> (MAB) . . . . .	20
2.2.1 Il problema MAB . . . . .	20
2.2.2 Una generalizzazione del problema MAB . . . . .	22
2.2.3 L'algoritmo <i>Thompson Sampling</i> . . . . .	23
2.3 L'algoritmo <i>Thompson Sampling</i> nella sorveglianza sequenziale . . . . .	26
<b>3 Casi reali di applicazione ed approccio proposto</b>	<b>33</b>
3.1 Due casi reali . . . . .	34
3.1.1 <i>Hotspot</i> nei processi di produzione additiva del metallo . . . . .	34
3.1.2 Il fenomeno dei brillamenti solari . . . . .	37
3.2 L'approccio proposto . . . . .	39
3.2.1 Identificazione delle mode . . . . .	40
3.2.2 Metodo di imputazione . . . . .	41
3.2.3 Strategia di campionamento . . . . .	44
<b>4 Applicazione dell'approccio proposto</b>	<b>47</b>
4.1 Identificazione degli <i>hotspot</i> . . . . .	47
4.1.1 Scelta dell'ampiezza di banda . . . . .	48
4.1.2 Scintille e loro riconoscimento . . . . .	50
4.1.3 Regioni luminose multiple . . . . .	52
4.1.4 Regione sparsa . . . . .	54
4.1.5 Sequenza di <i>frame</i> . . . . .	55
4.2 Identificazione dei brillamenti solari . . . . .	56
4.2.1 Imputazione e campionamento: 10% dei pixel totali . . . . .	56
4.2.2 Imputazione e campionamento: 2.95% dei pixel totali . . . . .	59

<b>Conclusioni</b>	<b>61</b>
<b>Appendice</b>	<b>63</b>
<b>Bibliografia</b>	<b>75</b>





# Introduzione

Negli ultimi decenni, con lo sviluppo delle tecnologie di acquisizione dei dati, la disponibilità di dati complessi e ad elevata dimensionalità è divenuta sempre più frequente e diffusa. Dall'ingegneria meccanica, alla genetica, all'astrofisica ed in molti altri contesti applicativi, l'enorme quantità di dati raccolti offre straordinarie opportunità per migliorare i sistemi di sorveglianza ed intraprendere tempestivamente azioni preventive e correttive, quando necessario.

Le immagini costituiscono una particolare tipologia di dati complessi e possono essere acquisite ad elevata frequenza tramite videocamere e sistemi di visione artificiale, dal costo notevolmente ridotto rispetto ad altri sensori impiegati per la sorveglianza in tempo reale di un processo o di un fenomeno dinamico. Tramite le sequenze di immagini acquisite ed i video registrati, viene catturata la dinamicità che caratterizza l'andamento di innovativi processi di produzione e l'evoluzione di numerosi fenomeni naturali. Tuttavia, difficoltà rilevanti emergono nell'analisi e nella sorveglianza in tempo reale di tali dati complessi e ad elevata dimensionalità, per i quali le tecniche di sorveglianza tradizionali, nella maggior parte dei casi, non sono direttamente applicabili.

In numerose applicazioni, inoltre, non è possibile acquisire tutte le misurazioni in ogni istante temporale. Con riferimento alla sorveglianza di flussi di immagini, in numerose applicazioni, ad esempio, non è possibile osservare la luminosità in corrispondenza di tutti i pixel che compongono ciascuna immagine acquisita, ma solo la luminosità in corrispondenza di un sottoinsieme di essi. In presenza di osservazioni parziali ad ogni istante temporale, le tecniche di sorveglianza tradizionali dovranno tener conto delle ulteriori difficoltà derivanti dalla presenza di dati mancanti. Sarà pertanto necessario implementare, in aggiunta ad un efficace schema di controllo, un metodo di imputazione

delle misurazioni mancanti ed una strategia di campionamento per selezionare, ad ogni istante temporale e in assenza di allarme, il sottoinsieme di pixel da osservare all'istante temporale successivo.

Nella presente trattazione si propone un nuovo approccio finalizzato alla ricostruzione e al campionamento sequenziale di flussi di dati da video o sequenze di immagini parzialmente osservabili. L'approccio è non parametrico e, per imputare le misurazioni mancanti e selezionare i pixel da osservare all'istante temporale successivo, prevede di identificare, in ogni istante temporale, quelle che si suppone siano le mode della distribuzione sottostante, parzialmente osservabile. Due ambiti applicativi estremamente diversi tra loro sono stati scelti volutamente per mostrare l'applicazione dell'approccio proposto.

La trattazione è strutturata come segue. Nel Capitolo 1 viene introdotta brevemente la sorveglianza statistica di processo e la sorveglianza sequenziale dei dati ad elevata dimensionalità, fornendo una panoramica delle recenti metodologie proposte in letteratura. Nella descrizione delle metodologie, sviluppate assumendo che siano disponibili solo osservazioni parziali in ogni istante temporale, si evidenziano le caratteristiche che saranno impiegate nell'approccio proposto. Nel Capitolo 2 viene descritto l'algoritmo *Thompson Sampling* ed alcune metodologie di sorveglianza che impiegano tale algoritmo come strategia di campionamento. L'approccio proposto, basato sull'identificazione delle mode, è illustrato nel Capitolo 3, assieme alla descrizione dei due casi reali. L'applicazione dell'approccio ai due casi reali è riportata nel Capitolo 4. Nelle Conclusioni sono riportate le considerazioni finali sui risultati raggiunti ed ulteriori contesti applicativi nei quali l'approccio può essere impiegato. L'Appendice contiene il codice R necessario per implementare l'intera procedura e riprodurre i risultati ottenuti.



# Capitolo 1

## La sorveglianza statistica dei dati ad elevata dimensionalità

### 1.1 Introduzione alla sorveglianza statistica

La qualità di un prodotto è un aspetto di fondamentale interesse per compagnie ed aziende di produzione che mirano a soddisfare le esigenze della clientela e a garantire un certo livello di conformità alle specifiche del prodotto finito. Al fine di garantire uno specifico livello di qualità del prodotto, è possibile valutare le sue proprietà e le sue caratteristiche dopo che il processo di produzione si è concluso. In alternativa, è possibile spostare l'attenzione dalla qualità del prodotto alla qualità del processo valutando, dunque, l'andamento e l'evoluzione del processo di produzione 'in linea'. La valutazione dell'andamento del processo in tempo reale, ossia la valutazione della fase di produzione del prodotto durante la sua esecuzione, consente di agire tempestivamente allo scopo di correggere e prevenire eventuali difetti ed anomalie del prodotto finito. L'obiettivo della valutazione e della sorveglianza del processo riguarda, pertanto, l'identificazione di eventuali andamenti indesiderati del processo che potrebbero compromettere la qualità dei prodotti. Quando s'intende sorvegliare in tempo reale (*online*) il processo, è cruciale identificare tempestivamente eventuali andamenti indesiderati al fine di interrompere, il più presto possibile, il processo di produzione qualora l'andamento del processo sia

instabile, indesiderato o, equivalentemente, fuori controllo (*Out-of-Control*, OC). Al contrario, quando il processo ha un andamento desiderato e opera in condizioni di stabilità, questo viene dichiarato in controllo (*In-Control*, IC).

Il termine 'sorveglianza statistica di processo' (*Statistical Process Monitoring*, SPM) si riferisce all'insieme di metodologie statistiche sviluppate allo scopo di sorvegliare l'andamento e l'evoluzione dei processi di produzione. Le metodologie statistiche sviluppate per la sorveglianza dei processi di produzione, oltre all'ambito industriale, sono largamente impiegate anche nei contesti legati alle scienze chimiche, fisiche, ambientali ed epidemiologiche quando s'intende sorvegliare in tali ambiti l'evoluzione di uno specifico fenomeno d'interesse. Per valutare l'evoluzione di un fenomeno o, in generale, l'andamento di un processo si disegna una carta di controllo per sorvegliare una o più caratteristiche di qualità che sono indice della condizione del processo. La carta di controllo è una rappresentazione grafica di una caratteristica di qualità che viene misurata o calcolata a partire da un campione di dati e, tale misurazione, viene rappresentata in funzione del numero del campione o del tempo. Si parla di sorveglianza statistica univariata quando si sorveglia un'unica caratteristica di qualità o un insieme di caratteristiche sorvegliate in modo indipendente l'una dall'altra mentre, si parla di sorveglianza statistica multivariata quando si sorveglia congiuntamente un insieme di caratteristiche di qualità. Si rimanda a Qiu (2014) e a Montgomery (2020) per ulteriori approfondimenti e dettagli sulle tecniche di sorveglianza per variabili univariate e multivariate.

Con le carte di controllo, rappresentando graficamente la statistica di controllo, ossia la misurazione della caratteristica di qualità, si evidenziano eventuali andamenti sospetti del processo e si evidenziano i campioni o gli istanti temporali in corrispondenza dei quali l'andamento del processo è fuori controllo. In questi casi viene segnalato un allarme e, successivamente, una fase di analisi ulteriore, chiamata diagnosi o analisi *post-segnale*, può essere condotta allo scopo di identificare la posizione spaziale, la forma e la dimensione dell'anomalia ed individuarne le cause.

Le metodologie statistiche sviluppate nell'ambito dell'SPM sono suddivise in due gruppi: Fase I e Fase II. In Fase I si analizza un insieme di dati di dimensione fissata, ordinati nel tempo e raccolti *offline*; il processo viene pertanto sorvegliato retrospettivamente. Purtroppo, in questa fase si ha a disposizione pochissima informazione sul

processo ed, in particolare, nella maggior parte delle applicazioni non è nota la distribuzione in controllo. L'obiettivo della Fase I è duplice: s'intende valutare la stabilità del processo e fornire una stima della distribuzione generatrice delle osservazioni in controllo. La performance delle metodologie di Fase I è strettamente legata alla corretta specificazione della distribuzione in controllo. Per questo motivo, la fase di stima della distribuzione in controllo, della sua forma, dei suoi parametri di posizione e di scala, dovrebbe essere condotta solo dopo aver verificato la stabilità del processo (Capizzi, 2015). A causa della limitata informazione relativa al processo che s'intende sorvegliare si consiglia di adottare un approccio non parametrico per sorvegliare retrospettivamente il processo. È cruciale sottolineare, infatti, che l'errata specificazione della distribuzione in controllo può determinare un grave deterioramento nella performance delle metodologie parametriche di Fase I. Incorrere in un'errata specificazione della distribuzione in controllo è uno scenario piuttosto verosimile poiché, spesso, in Fase I non si ha alcuna informazione riguardo la forma della distribuzione in controllo. Di conseguenza, sembra preferibile ricorrere ad un approccio non parametrico per la sorveglianza in Fase I; si veda per esempio Bell et al. (2014); Cheng & Shiau (2014); Capizzi & Masarotto (2017).

In Fase II il processo viene sorvegliato sequenzialmente (*online*) al fine di identificare il più rapidamente possibile eventuali deviazioni dall'andamento in controllo. A differenza della Fase I, in Fase II il campionamento non si arresta fino alla segnalazione di un allarme ed è estremamente rilevante la velocità con cui la metodologia di sorveglianza *online* identifica un andamento fuori controllo. L'interesse principale dell'analisi di Fase II riguarda la minimizzazione del ritardo con cui la metodologia di sorveglianza *online* segnala un allarme, inoltre, in Fase II si suppone di conoscere la distribuzione del processo in controllo, stimata con l'analisi di Fase I.

Le carte di controllo di Fase II possono essere suddivise in: (i) carte di tipo Shewhart (Shewhart, 1931) che impiegano esclusivamente i dati osservati all'istante temporale corrente per individuare un eventuale cambiamento indesiderato e (ii) carte CUSUM (*Cumulative Sum*) ed EWMA (*Exponentially Weighted Moving Average*), le quali impiegano tutti i dati acquisiti fino all'istante temporale corrente (Qiu, 2014).

Le metodologie di sorveglianza di Fase II per variabili multivariate potrebbero non essere adeguate in presenza di dati ad elevata dimensionalità. Ad esempio, sorvegliando

un processo tramite immagini o video, le dimensioni di ciascuna immagine acquisita in ogni istante temporale possono raggiungere l'ordine di grandezza di 1 milione di pixel (Yan et al., 2018). Inoltre, le immagini ad elevata risoluzione sono spesso acquisite ad elevata frequenza e sebbene la disponibilità di grandi moli di dati fornisca notevoli opportunità per la sorveglianza *online*, essa costituisce una sfida significativa per l'analisi in tempo reale e la diagnosi *post-segnale*.

La trattazione seguente si concentrerà sulla sorveglianza *online* di immagini e video che si traduce nell'analisi in tempo reale della luminosità associata a ciascun pixel. Nella maggior parte delle applicazioni reali, la variazione della luminosità registrata nel tempo cattura l'andamento del processo che è dinamico anche quando il processo è in controllo. Evidenti difficoltà compaiono nell'approssimazione della distribuzione IC a causa della non stazionarietà e della correlazione spazio-temporale che caratterizza i pixel o, equivalentemente, che caratterizza i flussi di dati (*data stream*) anche in condizioni di stabilità. Infatti, le misurazioni delle luminosità acquisite in corrispondenza di pixel adiacenti tra loro sono spazialmente correlate e le misurazioni acquisite nel tempo, in corrispondenza di ciascun pixel, sono temporalmente correlate. Di conseguenza, il calcolo del limite di controllo, ossia la soglia di riferimento con cui viene confrontata la statistica di controllo per segnalare o meno un allarme, può risultare difficoltoso da calcolare nella maggior parte delle applicazioni reali. Ad esempio, nella sorveglianza sequenziale di alcuni processi, come il *Selective Laser Melting* (SLM) che sarà descritto nel Capitolo 3, si può ricorrere ad approcci non convenzionali per definire una regola d'allarme adeguata in presenza di un andamento del processo che cambia continuamente anche in controllo (Colosimo & Grasso, 2018).

Quando è possibile acquisire tutte le misurazioni ad ogni istante temporale, le tecniche di sorveglianza analizzano la variazione della luminosità nel tempo, tuttavia, in numerose applicazioni non è possibile acquisire ed analizzare le luminosità registrate in corrispondenza di tutti i pixel che compongono ciascuna immagine. Ad esempio, per sorvegliare l'attività solare ed identificare i brillamenti solari vengono raccolte immagini (*frame*) ad alta risoluzione ed è possibile analizzare solo un ridotto numero di pixel in ogni istante temporale, a causa della limitata banda di trasmissione dei dati e delle limitate risorse computazionali per l'analisi in tempo reale. Quando è disponibile solo

un sottoinsieme di misurazioni in ogni istante temporale, le metodologie tradizionali di sorveglianza sequenziale devono tener conto di ulteriori difficoltà derivanti dalla presenza di dati mancanti. La trattazione seguente si concentrerà, in particolare, nel caso in cui siano disponibili solo osservazioni parziali per la sorveglianza in tempo reale del processo, condotta tramite immagini o video.

## 1.2 La sorveglianza sequenziale dei flussi di immagini

Con il progresso delle tecnologie di acquisizione dei dati, avvenuto negli ultimi decenni, la disponibilità di dati complessi e ad elevata dimensionalità è divenuta sempre più frequente e diffusa. L'introduzione di microchip, sensori avanzati, *imaging* digitale e comunicazioni *wireless* ha rivoluzionato numerosi processi di acquisizione dei dati nell'ambito della sorveglianza di processi. La disponibilità di grandi moli di dati offre, inoltre, enormi opportunità per la sorveglianza, la previsione, il miglioramento dell'affidabilità e la riduzione dei costi operativi (Hong et al., 2018).

Quando i dati ad elevata dimensionalità (*high-dimensional*, HD) sono accessibili per la sorveglianza sequenziale di un dato fenomeno è necessario condurre una fase preliminare di analisi al fine di pre-processare e sintetizzare i dati per condurre le analisi successive. Le immagini ed i video costituiscono un esempio di dati complessi e ad elevata dimensionalità. Il loro utilizzo nel contesto della sorveglianza si è notevolmente intensificato grazie allo sviluppo e alla diffusione dei sistemi di visione artificiale (*Machine Vision Systems*, MVS). Un sistema di visione artificiale è un sistema informatico che utilizza uno o più dispositivi di acquisizione immagini, come telecamere o dispositivi a raggi X, per fornire immagini da analizzare ed interpretare (Megahed et al., 2011). Le immagini acquisite sequenzialmente, spesso ad elevata frequenza, devono essere analizzate preliminarmente applicando, ad esempio, la riduzione del rumore, la compressione delle immagini stesse ed eventuali altre tipologie di elaborazione dei dati condotte necessariamente prima di effettuare la fase di sorveglianza del processo.

A causa della velocità di acquisizione dei dati, dell'enorme quantità e della complessità del dato stesso che deve essere necessariamente elaborato in tempo reale, senza

perdita d'informazione, le tecniche e le metodologie tradizionali di sorveglianza possono risultare inefficienti, richiedendo adattamenti e modifiche.

Nei paragrafi seguenti viene riportata una breve panoramica delle principali metodologie, sviluppate di recente, per la sorveglianza di flussi di immagini. Le metodologie esistenti in letteratura possono essere suddivise in due gruppi in base alla disponibilità di osservazioni da impiegare, in ogni istante temporale, per sorvegliare l'andamento del processo. In particolare, le metodologie citate nel paragrafo 1.2.1 sono state sviluppate assumendo che tutte le osservazioni siano disponibili in ogni istante temporale, ossia supponendo di poter osservare tutti i pixel che compongono ciascuna immagine acquisita. Diversamente, le metodologie citate nel paragrafo 1.2.2 sono state sviluppate assumendo che il processo generatore sia, in ogni istante temporale, parzialmente osservabile e solo un ridotto numero di pixel sia disponibile per sorvegliare il processo.

### **1.2.1 Sorveglianza e riduzione della dimensionalità**

Per sorvegliare un processo tramite flussi di immagini, la luminosità registrata in corrispondenza di ciascun pixel e la variazione della luminosità nel tempo viene analizzata sequenzialmente. Per fronteggiare le difficoltà derivanti dall'elevata dimensionalità dei flussi di dati, invece di sorvegliare la luminosità di ogni singolo pixel, che compone ciascuna immagine acquisita, è possibile ricorrere a svariate tecniche di riduzione della dimensionalità al fine di sorvegliare specifiche caratteristiche estratte dalle immagini. Ad esempio, è possibile sorvegliare indici sintetici di qualità oppure impiegare delle tecniche di riduzione della dimensionalità come l'analisi delle componenti principali (PCA) e sue estensioni.

Un insieme di immagini acquisite sequenzialmente può essere considerato come un tensore, il quale, tramite decomposizioni tensoriali, può essere ridotto rispetto alla dimensionalità iniziale, preservando il contenuto informativo. Un tensore può essere definito come un array multidimensionale; ad esempio, un insieme di 100 immagini, ciascuna avente dimensione  $M \times N$ , può essere considerato come un array di dimensione  $M \times N \times 100$ , ossia come un tensore di ordine 100. Per una panoramica riguardante i tensori di ordine superiore, le decomposizioni tensoriali e le loro applicazioni si rimanda

a Kolda & Bader (2009). Diversi metodi di decomposizione tensoriale sono stati sviluppati per ridurre la dimensionalità del tensore originale a una combinazione di tensori di rango inferiore; ne sono un esempio l'analisi delle componenti principali *unfold* (UPCA), l'analisi multilineare delle componenti principali (MPCA), l'analisi multilineare delle componenti principali incorrelate (UMPCA), la decomposizione *Candecomp/Parafac* (CP) e la decomposizione in tensori di rango uno (TROD). In Yan et al. (2015), scomponendo il tensore originale, tramite una tra le decomposizioni tensoriali proposte, si estraggono le caratteristiche rilevanti contenute nella sequenza acquisita di immagini e tali caratteristiche, assieme alla componente residua della decomposizione tensoriale, vengono sorvegliate tramite tecniche di sorveglianza multivariata (Qiu, 2014).

Le metodologie di sorveglianza sequenziale, basate su metodi di decomposizione tensoriale per la caratterizzazione della struttura di correlazione spazio-temporale dei dati, si pongono l'obiettivo di sorvegliare ed identificare eventuali variazioni del *pattern* della sequenza di immagini acquisite nel tempo, ossia variazioni delle strutture ricorrenti che si suppone rappresentino lo stato in controllo del processo. L'informazione spaziale relativa alla posizione dell'anomalia entro l'immagine non viene, purtroppo, sfruttata ed, in questo contesto, è spesso difficoltoso condurre la fase di analisi *post-segnale* per la diagnosi dell'anomalia identificata.

Altre metodologie di sorveglianza, al contrario, impiegano l'informazione spaziale contenuta nei dati allo scopo di identificare, entro l'immagine, la posizione dell'anomalia. Applicando tali metodologie, spesso denominate "carte di controllo spaziali", la statistica di controllo viene rappresentata in funzione dello spazio, non del tempo (Megahed et al., 2011). Le carte di controllo spaziali impiegano una finestra mobile che, spostandosi, percorre l'intera immagine. In corrispondenza di ciascuna finestra viene calcolata una statistica di controllo e, successivamente, le statistiche di controllo vengono rappresentate graficamente per identificare le porzioni dell'immagine in cui vi è un'anomalia (si veda per esempio Jiang & Jiang, 1998; Jiang et al., 2005; Lu & Tsai, 2005).

In assenza di immagini di riferimento (*gold standard*) che rappresentano lo stato in controllo del processo, le carte di controllo spaziali vengono impiegate per sorvegliare superfici testurizzate che si assume siano state generate da un processo stocastico

bidimensionale (Bui & Apley, 2018a,b).

Idealmente, nel contesto della sorveglianza in tempo reale di flussi di immagini, seppure vi sia l'esigenza di ridurre la dimensionalità dei dati, si vorrebbe incorporare nella metodologia di sorveglianza l'informazione sul processo sia spaziale che temporale, soprattutto quando il flusso di immagini acquisite cattura l'andamento del processo che risulta dinamico anche in controllo. La comparsa di deviazioni dallo stato in controllo si manifesta attraverso variazioni e perturbazioni di uno stato di per sé dinamico e, nella maggior parte dei casi, risulta difficoltoso approssimare le distribuzioni in controllo e fuori controllo del processo. Pertanto, le tecniche di sorveglianza tradizionali di Fase II dovranno necessariamente subire adattamenti e modifiche.

Nel contesto della sorveglianza in tempo reale di video, al fine di caratterizzare e sintetizzare la correlazione spazio-temporale che caratterizza la dinamica in controllo del processo, in Colosimo & Grasso (2018) viene proposta un'estensione dell'analisi delle componenti principali denominata *Spatially weighted T-mode PCA* (ST-PCA). La ST-PCA è una particolare estensione dell'analisi delle componenti principali *T-mode* (T-PCA, Jolliffe, 2002) che include una matrice di pesi spaziali nella decomposizione della matrice di varianze e covarianze allo scopo di descrivere la contiguità spaziale dei pixel in ciascun *frame* del video (Colosimo & Grasso, 2018). La ST-PCA viene applicata per la sorveglianza in linea dei processi di produzione additiva (AM) del metallo, al fine di identificare e localizzare eventuali difetti dei prodotti. Nei processi di AM, i prodotti vengono realizzati strato dopo strato e, con la metodologia proposta, s'intende rilevare le anomalie locali che potrebbero verificarsi durante la produzione di ciascuno strato del prodotto. L'obiettivo della sorveglianza riguarda, in particolare, l'identificazione di fenomeni di surriscaldamento locali, chiamati '*hot-spot*', causati da scambi di calore fuori controllo con il materiale circostante (Grasso et al., 2021). Una regione affetta da un *hotspot* rimane calda e, dunque, luminosa per un periodo di tempo più lungo ed è caratterizzata da un tasso di raffreddamento più lento rispetto alle zone di fusione regolare. Allo scopo di rilevare gli *hotspot* e segnalare un allarme, alla ST-PCA viene abbinata una regola di allarme basata sul raggruppamento individuato dall'algoritmo delle *k*-medie (Hennig et al., 2015).

Un ulteriore approccio alla sorveglianza in tempo reale di flussi di dati ad elevata



dimensionalità, che impiega un metodo innovativo per la riduzione della dimensionalità, viene proposto in Yan et al. (2018). La metodologia proposta prende il nome dalla decomposizione spazio-temporale (ST-SSD) impiegata per ridurre la dimensionalità dei dati, ed è ideata per la sorveglianza in tempo reale sia di profili singoli e multipli, sia di sequenze di immagini o video acquisiti con frequenza elevata, caratterizzati da un'evidente struttura di correlazione spazio-temporale e da un andamento non stazionario. A differenza dei metodi basati sull'analisi delle componenti principali e sue estensioni, lo schema ST-SSD è direttamente applicabile per la sorveglianza di *data stream* non stazionari. La tecnica di riduzione della dimensionalità impiegata decompone le misurazioni acquisite in ogni istante temporale e stima le caratteristiche anomale che vengono sorvegliate nel tempo. Lo schema di controllo ST-SSD, oltre a sorvegliare ed identificare l'istante temporale in cui il processo va fuori controllo, identifica la posizione spaziale dell'anomalia che rappresenta un'importante informazione per condurre un'opportuna analisi *post-segnale*. La metodologia ST-SSD viene applicata a tre casi reali, due dei quali si riferiscono alla sorveglianza in tempo reale di flussi di immagini; in particolare, la metodologia viene applicata (i) per l'identificazione dei difetti superficiali in un processo di laminazione dell'acciaio e (ii) per l'identificazione dei brillamenti solari che possono verificarsi sulla superficie del Sole.

### 1.2.2 Sorveglianza in presenza di osservazioni parziali

Le metodologie di sorveglianza sequenziale, citate nel paragrafo precedente, sono state sviluppate assumendo che il processo sia completamente osservabile, ossia assumendo di poter acquisire ed analizzare le luminosità registrate in corrispondenza di tutti i pixel che compongono ciascuna immagine. Le misurazioni della luminosità acquisite in ogni istante temporale  $t$ , ossia in corrispondenza di ciascuna immagine di dimensione  $M \times N$ , possono essere memorizzate in un vettore  $p$ -dimensionale  $\mathbf{X}_t \in \mathbb{R}^p$ , con  $p = MN$ .

Sfortunatamente, in numerose applicazioni, solo un sottoinsieme di  $m$  misurazioni è disponibile ad ogni tempo  $t$ , con  $m < p$ . Ad esempio, tale scenario si verifica quando, ad ogni istante temporale, solo un numero ridotto di sensori può essere attivato a causa della durata limitata delle batterie; oppure, quando solo le misurazioni di un sottoinsieme di

$m$  pixel sono disponibili a causa della limitata banda di trasmissione dei dati, delle limitate risorse *hardware* e limitate risorse computazionali per l'analisi in tempo reale.

In presenza di osservazioni parziali ad ogni istante temporale, le metodologie tradizionali di sorveglianza sequenziale devono tener conto di ulteriori difficoltà derivanti dalla presenza di dati mancanti. Oltre alla progettazione di uno schema di controllo efficace, le metodologie di sorveglianza sequenziale dovranno sviluppare: (i) un metodo di ricostruzione delle  $p - m$  misurazioni mancanti, sulla base delle quali, assieme alle  $m$  misurazioni osservate, verrà valutato lo stato del processo e (ii) una strategia di campionamento per selezionare, ad ogni istante temporale, un sottoinsieme di *data stream* da osservare quando lo schema non segnala un allarme.

La strategia di campionamento implementata dovrà selezionare, sequenzialmente e in maniera adattiva, il sottoinsieme di sensori o pixel in modo da esplorare e acquisire, anche al tempo successivo, le misurazioni che sono indice di un andamento sospetto al tempo corrente. La strategia di campionamento dovrà adottare e bilanciare dunque, ad ogni istante temporale, due principi opposti: (i) l'allocazione casuale dei *data stream* quando non vi è evidenza di un andamento sospetto (esplorazione) e (ii) l'allocazione dei *data stream* nelle regioni spaziali sospette (sfruttamento), quando vi è evidenza di una possibile deviazione dallo stato in controllo del processo. Quest'ultimo principio favorisce il cosiddetto sfruttamento delle risorse poiché, allocando i flussi di dati nelle regioni sospette, lo schema di controllo segnalerà l'eventuale andamento indesiderato con un ritardo inferiore. Il primo principio, al contrario, favorisce l'esplorazione delle risorse; ad esempio, nel contesto della sorveglianza sequenziale di flussi di immagini, questo principio favorirà l'esplorazione e l'osservazione di pixel distribuiti in maniera casuale entro il *frame* acquisito all'istante temporale corrente.

In letteratura, approcci parametrici e non parametrici sono stati proposti per sorvegliare sequenzialmente dati ad elevata dimensionalità in presenza di osservazioni parziali. Invece di sorvegliare i dati originali è possibile, in alcuni casi, stimare un adeguato modello spazio-temporale e sorvegliare i residui del modello, assumendo che questi siano indipendenti ed identicamente distribuiti e abbiano distribuzione nota (Wang & Mei, 2013). Sorvegliando flussi di dati indipendenti e identicamente distribuiti è possibile impiegare tecniche di sorveglianza parametriche (Qiu, 2014); tuttavia, stimare un modello

spazio-temporale in presenza di *data stream* non stazionari che descrivono l'andamento dinamico del processo e ricavare, in tempo reale, i residui del modello da sorvegliare può risultare difficoltoso nella pratica. Pertanto, l'assunzione di indipendenza ed identica distribuzione dei flussi di dati che s'intende sorvegliare può risultare irrealistica nella maggior parte delle applicazioni reali. Per rilassare le assunzioni distributive, approcci non parametrici sono stati sviluppati generalizzando carte di controllo, disegnate assumendo che il processo sia completamente osservabile, al caso in cui siano presenti solo osservazioni parziali del processo in ogni istante temporale. In presenza di osservazioni parziali e, dunque, in presenza di dati mancanti in ogni istante temporale, sono stati sviluppati metodi di compensazione delle misurazioni mancanti che impiegano un parametro di compensazione fissato a priori o, in alternativa, che imputano le misurazioni mancanti tramite la *data augmentation*, ossia tramite l'aumento dei dati che avviene sfruttando tutta l'informazione disponibile. Tuttavia, poiché flussi di dati adiacenti sono tra loro spazialmente correlati, anche l'informazione spaziale può essere impiegata per imputare le misurazioni mancanti, sulla base delle misurazioni acquisite in corrispondenza dei flussi di dati osservati nelle vicinanze. Inoltre, l'informazione spaziale può essere impiegata anche nella fase di campionamento poiché è cruciale selezionare *data stream* nelle regioni spaziali sospette, allo scopo di identificare un eventuale cambiamento indesiderato col minor ritardo possibile.

Tra gli approcci proposti per la sorveglianza *online* in presenza di osservazioni parziali, Liu et al. (2015) sviluppano uno schema di controllo denominato TRAS (*Top-R based Adaptive Sampling*) per sorvegliare sequenzialmente *data stream* indipendenti ed identicamente distribuiti, aventi distribuzione normale. La procedura TRAS prevede l'assegnazione di due statistiche locali di tipo CUSUM in corrispondenza di ciascun *data stream*. In corrispondenza dei *data stream* non osservati, le statistiche locali vengono aggiornate impiegando un parametro di compensazione fissato a priori. Le statistiche locali vengono ordinate in maniera decrescente e le  $R$  statistiche locali più grandi vengono aggregate in una somma (*Top-R rule*, Mei, 2011) che viene confrontata con il limite di controllo. Qualora lo schema non segnali un allarme, i *data stream* corrispondenti alle  $m$  statistiche locali più grandi (con  $m < p$ ) verranno campionati al tempo successivo. Per rilassare l'assunzione di normalità e sviluppare uno schema di controllo

per la sorveglianza di flussi di dati scambiabili, in Xian et al. (2018) viene proposta la metodologia NAS, una strategia di campionamento adattiva e non parametrica basata sull'anti-rango. Per affrontare il problema dei dati mancanti si impiega, come per la TRAS, un parametro di compensazione fissato a priori, generalizzando, al caso in cui il processo sia parzialmente osservabile, la procedura CUSUM proposta in Qiu & Hawkins (2001, 2003) basata sull'indicatore anti-rango. In assenza di allarme, per selezionare i *data stream* da osservare al tempo successivo, similmente alla procedura TRAS, NAS prevede che le singole statistiche locali calcolate in corrispondenza di ciascun *data stream* vengano ordinate e i *data stream* corrispondenti alle  $m$  statistiche più grandi vengano campionati al tempo successivo. La reattività degli schemi di controllo TRAS e NAS, nell'identificare una deviazione dallo stato in controllo è, purtroppo, fortemente influenzata dalla scelta del parametro di compensazione fissato a priori. Inoltre, non è esplicita la relazione che lega la reattività degli schemi di controllo con il valore del parametro di compensazione, pertanto, la scelta del parametro può risultare critica.

In alternativa alle formulazioni che prevedono l'impiego di un parametro di compensazione fissato a priori, Xian et al. (2021) propongono una strategia di campionamento, basata sull'anti-rango, che impiega la *data augmentation* e denominata R-SADA. Con la *data augmentation*, l'imputazione delle osservazioni mancanti avviene in maniera dinamica sfruttando le misurazioni provenienti dai *data stream* osservati all'istante temporale corrente e, con essa si ricava un vettore di misurazioni che viene impiegato nella fase di sorveglianza del processo, la quale adotta la procedura CUSUM proposta in Qiu & Hawkins (2001). La strategia R-SADA è parametrica poiché impiega la funzione di densità di probabilità e la funzione di ripartizione del processo IC nella fase di *data augmentation*, le quali si assume siano note da dati di Fase I. Tuttavia, richiedere che la distribuzione del processo IC sia nota prima di condurre la fase di sorveglianza è un requisito difficile da soddisfare nella pratica poiché, spesso, non si ha a disposizione un campione sufficientemente numeroso e rappresentativo del processo IC.

Le strategie di campionamento impiegate in TRAS, NAS ed R-SADA per selezionare il sottoinsieme di *data stream* da osservare, valutano la grandezza delle specifiche statistiche locali, calcolate in ogni istante temporale in ciascuno dei tre schemi di controllo.

In questo modo, l'allocazione effettuata dalle tre strategie non bilancia entrambi i principi dello sfruttamento e dell'esplorazione, ma seleziona i flussi di dati adottando il solo principio dello sfruttamento delle risorse. Inoltre, nell'allocazione adattiva dei flussi di dati, le tre strategie di campionamento non sfruttano l'informazione relativa alla vicinanza spaziale dei flussi di dati. Nella sorveglianza sequenziale di flussi di immagini, è ragionevole ipotizzare che l'occorrenza di un'anomalia influenzi la luminosità registrata in corrispondenza di uno o più gruppi di pixel adiacenti tra loro. Una metodologia di sorveglianza ed una strategia di campionamento adattiva che incorpora l'informazione relativa alla posizione spaziale dei flussi di dati è proposta in Wang et al. (2018). La procedura proposta, denominata SASAM (*Spatial-Adaptive Sampling And Monitoring*), in assenza di allarme seleziona i *data stream* da osservare al tempo successivo suddividendo in due sottogruppi gli  $m$  *data stream* osservabili, in modo da soddisfare entrambi i principi di esplorazione e sfruttamento delle risorse. In corrispondenza di ciascun flusso di dati, tra i  $p$  totali, viene calcolata e aggiornata ad ogni istante temporale una statistica di controllo locale. Per risolvere il problema delle osservazioni mancanti, l'informazione spaziale viene sfruttata e vengono aggiornate anche le statistiche locali relative ai  $p - m$  flussi di dati non osservati al tempo corrente, sfruttando le misurazioni acquisite nelle vicinanze. Se la statistica locale avente il valore massimo tra tutte non eccede il limite di controllo, lo schema non segnala un allarme e si campionano  $m$  flussi di dati allocando un sottogruppo di essi attorno al flusso al quale è associata la statistica locale massima, e i rimanenti in maniera casuale. La strategia di campionamento impiegata con SASAM è adattiva poiché le numerosità dei due sottogruppi variano al variare del valore assunto dalla statistica locale massima; tuttavia, la strategia è in grado di campionare i flussi di dati che si concentrano in un'unica regione spaziale sospetta. In presenza di due o più regioni in cui l'andamento del processo diviene sospetto e indesiderato, la strategia di campionamento non è in grado di identificare il numero corretto di regioni fuori controllo e di campionare simultaneamente in ognuna di esse. Inoltre, con la procedura SASAM si assume che, quando il processo è in controllo, i flussi di dati siano indipendenti e identicamente distribuiti, con distribuzione normale. Le metodologie citate si basano su assunti distributivi che possono risultare particolarmente irrealistici nella pratica, inoltre, nessuna di queste preserva la struttura di correlazione spazio-temporale

che caratterizza i flussi di dati anche quando il processo è in controllo.

Al momento, in letteratura, è stata proposta un'unica metodologia per la sorveglianza *online* di flussi di dati HD in presenza di osservazioni parziali che non richiede siano soddisfatti particolari assunti distributivi; come, ad esempio, l'assunzione di normalità dei flussi di dati. La metodologia proposta in Gómez et al. (2022), denominata TSS (*Tensor Sequential Sampling*), è infatti l'unica che sorveglia i dati nel loro formato originale e impiega, per la costruzione dello schema, la correlazione spazio-temporale naturalmente presente nei dati. I flussi di dati, acquisiti in tempo reale, vengono considerati sotto forma di tensore incompleto il quale viene decomposto in tre componenti. Con la decomposizione e, in particolare, con la stima della prima componente si provvede all'imputazione delle misurazioni mancanti mentre, la seconda componente stimata viene impiegata nella fase di sorveglianza del processo. Qualora sia avvenuto un cambiamento indesiderato o, equivalentemente, il processo sia fuori controllo, gli elementi della seconda componente saranno più elevati rispetto al caso in cui il processo sia in controllo. Pertanto, la somma degli elementi della seconda componente, presi in valore assoluto, viene sorvegliata tramite una statistica di controllo di tipo EWMA (Qiu, 2014). Se lo schema segnala un allarme, si effettua una fase di diagnosi *post-segnale* che ha lo scopo di identificare la regione spaziale affetta dal cambiamento indesiderato. In assenza di allarme, si impiega una strategia di campionamento adattiva per selezionare i *data stream* da osservare al tempo successivo, bilanciando i principi di esplorazione e sfruttamento delle risorse. In particolare, per attribuire ad ogni *data stream* una certa probabilità di essere selezionato al tempo successivo, si impiega una mistura di due funzioni di massa di probabilità, ciascuna delle quali favorisce uno dei due principi. Seppure la metodologia TSS sia l'unica che sorveglia i dati nel loro formato originale senza richiedere che siano soddisfatte particolari assunzioni distributive, l'intera procedura richiede un tempo computazionale più elevato rispetto alle altre metodologie proposte in letteratura. In Gómez et al. (2022) si riporta, inoltre, l'applicazione di TSS a due casi reali, entrambi riferiti alla sorveglianza *online* di flussi di dati in presenza di osservazioni parziali. Uno dei due casi reali si riferisce alla sorveglianza di flussi di immagini e, in particolare, alla sorveglianza *online* dell'attività solare allo scopo di identificare i brillamenti che possono verificarsi sulla superficie del Sole. Il fenomeno dei brillamenti

solari viene descritto, in dettaglio, nel Capitolo 3. La sequenza di immagini raccolta allo scopo di sorvegliare l'attività solare viene impiegata anche in Liu et al. (2015), Xian et al. (2021) e Wang et al. (2018) per valutare le performance degli schemi di controllo, rispettivamente TRAS, R-SADA e SASAM, applicati al medesimo caso di studio.

Per identificare, il più presto possibile, un eventuale andamento indesiderato del processo e segnalare tempestivamente un allarme, è cruciale che la strategia di campionamento implementata selezioni il sottoinsieme di pixel da osservare, bilanciando entrambi i principi dell'esplorazione e dello sfruttamento delle risorse. Numerose strategie di campionamento affrontano il compromesso tra l'esplorazione e lo sfruttamento delle risorse, in un contesto di risorse limitate. Tra queste, l'algoritmo *Thompson Sampling* ha ricevuto, negli ultimi anni, notevole attenzione ed è stato impiegato in alcune metodologie di sorveglianza sequenziale. Tali metodologie assieme all'algoritmo sono descritte, di seguito, nel Capitolo 2.





## Capitolo 2

# Sorveglianza sequenziale e strategie di campionamento

### 2.1 Generalità sulle strategie di campionamento

In presenza di osservazioni parziali le metodologie di sorveglianza sequenziale devono necessariamente sviluppare, oltre allo schema di controllo, una strategia di campionamento per selezionare ad ogni istante temporale, in assenza di allarme, un sottoinsieme di *data stream* da osservare al tempo successivo. In particolare, come anticipato, è d'interesse selezionare il sottoinsieme di *data stream* da osservare, bilanciando il compromesso tra l'esplorazione e lo sfruttamento delle risorse in presenza di risorse limitate.

Il problema di scegliere sequenzialmente il sottoinsieme di  $m$  *data stream* da osservare è riconducibile al problema del *Combinatorial Multi-Armed Bandit* (CMAB), in cui un giocatore d'azzardo deve scegliere sequenzialmente un sottoinsieme di  $m$  leve da azionare, tra  $p$  leve di cui è dotata una *slot machine*, con l'obiettivo di massimizzare la vincita finale. La formulazione più semplice del problema CMAB, prevede che il giocatore possa azionare ad ogni istante temporale una sola leva ( $m = 1$ ) tra le  $p$  disponibili e, in questo caso, il problema viene chiamato *Multi-Armed Bandit* (MAB). Alcune strategie di campionamento sono state sviluppate per risolvere il problema MAB e le sue generalizzazioni; ne sono un esempio l'*Upper Confidence Bound* (UCB) proposto in Lai & Robbins (1985) ed il *Thompson Sampling* (TS) proposto, per la prima volta in letteratura, in Thompson (1933). Queste strategie, in grado di bilanciare il compromesso tra

l'esplorazione e lo sfruttamento delle risorse, possono essere impiegate in maniera efficace come strategie di campionamento adattive nel contesto della sorveglianza sequenziale in presenza di osservazioni parziali. In particolare, TS è una strategia di campionamento semplice da implementare e, a differenza di UCB, è applicabile al contesto della sorveglianza sequenziale in presenza di osservazioni parziali, anche quando i flussi di dati sono ad elevata dimensionalità (Guo et al., 2022).

## 2.2 *Thompson Sampling per il problema del Multi-Armed Bandit (MAB)*

Il campionamento di Thompson, di seguito introdotto come algoritmo TS (Thompson, 1933), è una strategia di campionamento impiegata per la soluzione in tempo reale di problemi decisionali, che sceglie sequenzialmente le azioni da compiere in modo da soddisfare un certo criterio prefissato. In particolare, impiegando un approccio bayesiano, l'algoritmo sceglie, in maniera automatica, la sequenza di azioni in grado di sfruttare al meglio i risultati ottenuti con le azioni scelte in precedenza, così massimizza la performance corrente e allo stesso tempo esplora le azioni che potrebbero rivelarsi più vantaggiose in futuro. L'algoritmo TS è dunque stato sviluppato per bilanciare in modo efficiente l'esplorazione e lo sfruttamento di tutte le possibilità in presenza di risorse limitate.

### 2.2.1 Il problema MAB

Con il problema del *Multi-Armed Bandit* (MAB) si fa riferimento all'allocazione sequenziale di un insieme limitato di risorse in modo da massimizzare il guadagno atteso (Robbins, 1952). In ogni istante temporale, scegliendo tra tutte le possibili azioni, se ne compie una e, in corrispondenza di ciascuna azione compiuta, si può osservare un successo oppure un fallimento. L'obiettivo è quello di selezionare nel tempo la sequenza di azioni con cui si ottiene la vincita massima. Inizialmente, tuttavia, non è nota la probabilità di successo associata a ciascuna azione, tra le  $p$  azioni disponibili ad ogni istante temporale. Pertanto, il problema si traduce nell'apprendimento delle probabilità

di successo associate a ciascuna azione, sulla base delle azioni compiute in precedenza e dei successi ottenuti.

Il termine MAB proviene dal contesto del gioco d'azzardo, in cui il termine "bandito" si riferisce alla *slot machine* che tende a far perdere il denaro ai giocatori (Russo et al., 2018). La *slot machine* è dotata di  $p$  leve. In ogni istante temporale, quando il giocatore aziona una leva, tra le  $p$  disponibili, può ricevere una vincita e la vincita restituita è indipendente dalle vincite restituite in passato dalla *slot machine* con la stessa leva ma anche da quelle che avrebbe restituito, scegliendo una leva diversa. La distribuzione della vincita associata a ciascuna leva non è nota e il giocatore può apprendere le probabilità di successo associate alle  $p$  leve solo continuando a sperimentare. In particolare, azionando di volta in volta una leva tra le  $p$  disponibili, il giocatore deve bilanciare il compromesso tra lo sfruttamento della conoscenza acquisita, e l'esplorazione delle leve non ancora azionate. Nell'immediato futuro, infatti, egli si aspetta di ricevere una vincita maggiore sfruttando le leve che in passato si sono rivelate più redditizie ma vorrebbe continuare anche ad esplorare le altre leve che potrebbero rivelarsi ancor più redditizie in futuro. Per bilanciare questi due comportamenti il giocatore può impiegare una strategia di campionamento per scegliere sequenzialmente le leve da azionare, allo scopo di massimizzare la vincita finale.

Il problema MAB è, pertanto, ricorrente in svariati contesti applicativi nei quali è necessario bilanciare il compromesso esistente tra l'esplorazione e lo sfruttamento delle risorse disponibili, continuando ad apprendere tramite la sperimentazione e l'interazione con il processo generatore. Ne sono un esempio i sistemi di raccomandazione (*recommender systems*) che devono suggerire prodotti o contenuti agli utenti al fine di massimizzare il loro interesse e la loro interazione con il sistema; oppure nell'ambito delle sperimentazioni cliniche la formulazione del problema MAB è ricorrente quando s'intende assegnare differenti terapie o trattamenti ad un gruppo di pazienti, ripetutamente e in qualsiasi ordine (Gittins, 1989). L'assegnazione delle terapie ai pazienti dovrà avvenire sequenzialmente e l'obiettivo della sperimentazione è quello di identificare la terapia più efficace tra quelle proposte.

### 2.2.2 Una generalizzazione del problema MAB

Il problema MAB può essere generalizzato al caso in cui, ad ogni istante temporale, il giocatore d'azzardo possa azionare simultaneamente  $m$  leve tra le  $p$  disponibili con  $1 < m < p$ . Il giocatore dovrà selezionare, ad ogni istante temporale, un sottoinsieme di  $m$  leve tra tutti i  $\binom{p}{m}$  possibili sottoinsiemi, in modo da massimizzare la vincita finale (CMAB). In particolare, ogni sottoinsieme di numerosità  $m$ , ottenibile a partire dalle  $p$  leve, è chiamato "super leva" e l'obiettivo, in ogni istante temporale, è quello di selezionare la super leva che massimizza la vincita al tempo corrente, bilanciando il compromesso tra l'esplorazione e lo sfruttamento (Zhang & Hoi, 2019). Si noti che la sequenza di azioni che massimizza la vincita è equivalente alla sequenza di azioni che minimizza la perdita (Hoi et al., 2021). Pertanto, è possibile riferirsi alla sequenza ottimale di azioni come alla sequenza in corrispondenza della quale si ottiene la vincita massima o, equivalentemente, la perdita minima. La funzione di vincita che s'intende massimizzare ad ogni istante temporale può essere lineare o non lineare rispetto alla vincita ottenuta da ciascuna leva, appartenente al sottoinsieme considerato. Inoltre, il sottoinsieme che massimizza la vincita all'istante corrente può non essere unico (Zhang & Hoi, 2019). Quando la funzione di vincita, chiamata anche funzione di ricompensa, non è lineare è necessario valutare tutti i  $\binom{p}{m}$  possibili sottoinsiemi di  $m$  leve ottenibili a partire dalle  $p$  leve disponibili (Guo et al., 2022). Quando il numero di leve disponibili è elevato e la funzione di vincita è non lineare, enumerare tutti i possibili sottoinsiemi di leve e valutare, per ciascuno di essi, la funzione di vincita può risultare estremamente oneroso dal punto di vista computazionale. Pertanto, nelle applicazioni reali si ricorre spesso ad una specificazione alternativa della funzione di vincita, in modo che essa sia lineare rispetto alla vincita ottenuta in corrispondenza di ciascuna leva appartenente al sottoinsieme considerato. In questo modo non è necessario enumerare tutti i possibili sottoinsiemi ed il costo computazionale è notevolmente ridotto. Ad ogni istante temporale si calcola la funzione di vincita in corrispondenza di ciascuna leva tra le  $p$  disponibili e il sottoinsieme selezionato è costituito dalle  $m$  leve in corrispondenza delle quali si osservano gli  $m$  valori più grandi, tra i  $p$  valori calcolati (Guo et al., 2022).

Tra gli algoritmi sviluppati per risolvere il problema MAB e CMAB, negli ultimi anni, l'algoritmo TS ha ricevuto notevole attenzione in quanto si tratta di un algoritmo

semplice da implementare che seleziona automaticamente, in ogni istante temporale, ciascuna leva da azionare in base alla sua probabilità di essere ottimale (Chapelle & Li, 2011).

### 2.2.3 L'algoritmo *Thompson Sampling*

Con riferimento alla formulazione del problema MAB nel contesto del gioco d'azzardo, si consideri una *slot machine* con  $p$  leve e si assuma che, ad ogni istante temporale, il giocatore d'azzardo possa selezionare una leva ( $m = 1$ ) tra le  $p$  disponibili. Dopo aver azionato al tempo  $t$  la leva  $k$ -esima, con  $k = 1, \dots, p$ , si assuma che il giocatore possa ricevere una vincita  $r_t$  con una certa probabilità  $\theta_k$ , non nota a priori, e che le vincite restituite dalla *slot machine* siano indipendenti.

La formulazione generale dell'algoritmo TS prevede che, inizialmente, il giocatore supponga che le probabilità di successo  $\theta = (\theta_1, \dots, \theta_p)$  siano distribuite secondo una distribuzione a priori congiunta. Al primo istante temporale ( $t = 1$ ), egli campiona casualmente  $p$  valori dalla distribuzione a priori congiunta o, equivalentemente, per l'indipendenza delle probabilità di successo il giocatore campiona casualmente un valore  $\hat{\theta}_k$  da ciascuna distribuzione a priori marginale. Successivamente, il giocatore aziona la leva  $j$ -esima che massimizza la vincita attesa, condizionatamente alla leva selezionata, ossia aziona la leva in corrispondenza della quale è stato osservato il valore massimo tra i  $p$  valori campionati  $\hat{\theta}_1, \dots, \hat{\theta}_p$ . Dopo aver azionato la leva la *slot machine* restituisce al giocatore la vincita  $r_1$  generata da una distribuzione di probabilità che dipende dai parametri non noti  $\theta$ . Sulla base della vincita restituita  $r_1$  la distribuzione a priori viene aggiornata e la procedura viene ripetuta al tempo successivo, aggiornando di volta in volta la distribuzione delle probabilità di successo  $\theta$  (Russo et al., 2018).

Una formulazione comune del problema MAB prevede che, dopo aver azionato la leva  $k$ -esima, la *slot machine* possa restituire al giocatore una vincita pari a 1, con probabilità  $\theta_k$  oppure una vincita pari a 0, con probabilità  $1 - \theta_k$ . In questo caso, il problema è chiamato *Bernoulli* MAB poiché la vincita associata a ciascuna leva è distribuita come una variabile casuale di Bernoulli (Russo et al., 2018). Le probabilità di successo  $\theta_1, \dots, \theta_p$  associate alle  $p$  leve disponibili sono fissate nel tempo ma non sono note a priori dal giocatore. Inizialmente, il giocatore suppone che ciascuna probabilità di

successo  $\theta_k$ , con  $k = 1, \dots, p$ , abbia distribuzione Beta con iperparametri  $\alpha_k > 0$ ,  $\beta_k > 0$ . La funzione di densità di probabilità a priori di  $\theta_k$  per  $k = 1, \dots, p$  è data da

$$p(\theta_k) = \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \theta_k^{\alpha_k-1} (1 - \theta_k)^{\beta_k-1} \quad (2.1)$$

dove  $\Gamma(\cdot)$  è la funzione gamma. Si noti che nel caso in cui  $\alpha_k = \beta_k = 1$  la distribuzione a priori in (2.1) coincide con la distribuzione uniforme nell'intervallo  $[0, 1]$ . La distribuzione a posteriori di  $\theta_k$  viene calcolata applicando la regola di Bayes. La scelta della distribuzione a priori Beta è particolarmente conveniente poiché la Beta è la distribuzione a priori coniugata per la Bernoulli. Pertanto, la distribuzione a posteriori di  $\theta_k$  sarà ancora una Beta con parametri aggiornati in base alla vincita ottenuta.

Dati i vettori di iperparametri  $\alpha = (\alpha_1, \dots, \alpha_p)$  e  $\beta = (\beta_1, \dots, \beta_p)$ , al primo istante temporale  $t = 1$ , l'algoritmo TS prevede di campionare casualmente un valore  $\hat{\theta}_k$  dalla distribuzione  $Beta(\alpha_k, \beta_k)$ , per ogni  $k = 1, \dots, p$ . Il giocatore azionerà quindi la  $j$ -esima leva in corrispondenza della quale è stato osservato il valore massimo tra i  $p$  valori campionati  $\hat{\theta}_1, \dots, \hat{\theta}_p$ . Pertanto, la scelta  $a_1$  effettuata dal giocatore al tempo  $t = 1$  corrisponde all'azionamento della leva  $j$ -esima

$$j = \underset{k}{\operatorname{argmax}} \hat{\theta}_k \rightarrow a_1 = j \quad (2.2)$$

Dopo aver azionato la leva  $j$ -esima, al tempo  $t = 1$ , la *slot machine* restituisce al giocatore una vincita  $r_1$  con  $r_1 \in \{0, 1\}$ . Dopo aver osservato il valore della vincita, gli iperparametri  $(\alpha_{a_1}, \beta_{a_1})$  vengono aggiornati come segue

$$(\alpha_{a_1}, \beta_{a_1}) \leftarrow (\alpha_{a_1} + r_1, \beta_{a_1} + 1 - r_1) \quad (2.3)$$

Al tempo  $t = 2$  si ripete la procedura campionando dalla distribuzione Beta con parametri aggiornati in corrispondenza della leva  $j$ -esima e dalle distribuzioni a priori per le altre leve, con  $k \neq j$ .

La procedura viene ripetuta ad ogni istante temporale e, di volta in volta, le distribuzioni delle probabilità di successo vengono aggiornate in base alle vincite ottenute nel tempo. Applicando la procedura e, dunque, continuando a sperimentare, il giocatore

---

**Algorithm 1** L'algoritmo TS per il problema MAB
 

---

**Require:**  $\alpha = (\alpha_1, \dots, \alpha_p)$ ,  $\beta = (\beta_1, \dots, \beta_p)$  vettori di iperparametri

**for**  $t = 1, \dots, T$  **do**

**for**  $k = 1, \dots, p$  **do**

        campionare  $\hat{\theta}_k \sim \text{Beta}(\alpha_k, \beta_k)$

**end for**

    scegliere  $a_t = \underset{k}{\operatorname{argmax}} \hat{\theta}_k$  e ricevere  $r_t$ , con  $r_t \in \{0, 1\}$

    aggiornare  $(\alpha_{a_t}, \beta_{a_t}) \leftarrow (\alpha_{a_t} + r_t, \beta_{a_t} + 1 - r_t)$

**end for**

---

è in grado di aggiornare sequenzialmente la sua conoscenza riguardo le probabilità di successo aggiornando le distribuzioni Beta associate a ciascun  $\theta_k$ , con  $k = 1, \dots, p$ , che divengono più concentrate attorno al vero valore della probabilità di successo.

In questo modo, allo scopo di massimizzare la vincita finale, l'algoritmo TS fornisce al giocatore la strategia per selezionare ad ogni istante temporale la leva da azionare, bilanciando il compromesso tra l'esplorazione e lo sfruttamento delle leve. L'algoritmo TS per il problema del *Bernoulli* MAB è presentato nell'Algoritmo 1.

L'algoritmo può essere generalizzato al caso in cui la vincita  $r_t$  restituita dalla *slot machine* ad ogni istante temporale non sia binaria. In particolare, nel problema più generale si assume che ciascuna vincita  $r_t$  sia stata generata da una distribuzione di probabilità arbitraria e non nota con supporto  $[0, 1]$ . L'Algoritmo 1 viene pertanto generalizzato inserendo un esperimento di Bernoulli con probabilità di successo pari a  $r_t$  e conducendo tale esperimento prima di aggiornare i parametri della distribuzione Beta (Agrawal & Goyal, 2012).

Nel problema CMAB, quando la funzione di vincita è lineare, l'algoritmo TS viene generalizzato ordinando in maniera decrescente il vettore dei valori campionati casualmente  $\hat{\theta}_{(1)}, \dots, \hat{\theta}_{(p)}$  con  $\hat{\theta}_{(1)} \geq \hat{\theta}_{(2)} \geq \dots \geq \hat{\theta}_{(p)}$ . In seguito, il giocatore aziona le  $m$  leve in corrispondenza delle quali si osservano i valori più grandi  $\hat{\theta}_{(1)}, \dots, \hat{\theta}_{(m)}$  (Zhang & Mei, 2022; Guo et al., 2022; Ye et al., 2023).

## 2.3 L'algoritmo *Thompson Sampling* nella sorveglianza sequenziale

Nel problema CMAB il giocatore seleziona sequenzialmente un sottoinsieme di  $m$  leve, con  $m > 1$ , al fine di massimizzare la vincita finale. Il problema CMAB, descritto nel paragrafo 2.2.2, può essere associato al caso in cui s'intende sorvegliare sequenzialmente un processo in presenza di osservazioni parziali e una strategia di campionamento viene impiegata allo scopo di selezionare, in assenza di allarme, un sottoinsieme di  $m$  *data stream* da osservare. In particolare, per il giocatore d'azzardo è d'interesse azionare sequenzialmente il sottoinsieme di leve alle quali sono associate le probabilità di successo più elevate. Similmente, nel contesto della sorveglianza sequenziale in presenza di osservazioni parziali, è d'interesse selezionare il sottoinsieme di *data stream* che, con elevata probabilità, ha sperimentato un cambiamento indesiderato all'istante temporale corrente (Guo et al., 2022).

A differenza delle formulazioni del problema MAB, riportate nel paragrafo 2.2, nel contesto della sorveglianza sequenziale le probabilità  $\theta_1, \dots, \theta_p$  associate a ciascun *data stream*, sono variabili nel tempo. La variabilità nel tempo della probabilità di sperimentare un cambiamento, associata a ciascun *data stream*, caratterizza qualsiasi applicazione nel contesto della sorveglianza sequenziale. In presenza di probabilità  $\theta_1, \dots, \theta_p$  variabili nel tempo il processo generatore o, equivalentemente, la *slot machine* che il giocatore d'azzardo aziona, è non stazionario (Russo et al., 2018). In questi casi non si smette mai di allocare risorse sia in favore dello sfruttamento della conoscenza acquisita, sia in favore dell'esplorazione, diversamente dal caso stazionario in cui, al passare del tempo e continuando a sperimentare, il numero di risorse allocate in favore dell'esplorazione tende a zero.

Nel contesto della sorveglianza sequenziale in presenza di osservazioni parziali, l'algoritmo TS può essere impiegato in maniera efficace come strategia di campionamento adattiva che seleziona, ad ogni istante temporale, sia i flussi di dati che mostrano andamenti sospetti nel tempo (sfruttamento), sia flussi di dati campionati casualmente (esplorazione) al fine di identificare, il più presto possibile, ulteriori andamenti indesiderati.



Recentemente sono state sviluppate alcune metodologie per la sorveglianza sequenziale di flussi di dati HD in presenza di osservazioni parziali che impiegano l'algoritmo TS come strategia di campionamento adattiva. Zhang & Mei (2022) sviluppano una metodologia che combina l'algoritmo TS con la procedura di Shiryaev-Roberts-Pollak per la sorveglianza sequenziale del processo (TSSRP). Per valutare lo stato del processo, si calcolano due tipologie di statistiche locali in corrispondenza di ciascun flusso di dati, tra i  $p$  flussi disponibili. Lo schema segnala un allarme quando la somma, calcolata in ogni istante temporale, delle  $R$  statistiche locali più grandi della prima tipologia eccede il limite di controllo (*Top-R rule*, Mei, 2011). In particolare, la statistica locale della prima tipologia è una generalizzazione della statistica di Shiryaev-Roberts ed indica quanto sia probabile che si sia verificato un cambiamento locale. Si rimanda a Tartakovsky et al. (2014) per approfondimenti sulla statistica di Shiryaev-Roberts. La statistica locale della seconda tipologia è legata al numero di campioni acquisiti nel tempo in corrispondenza di ogni flusso di dati. Per ogni flusso di dati, osservato all'istante temporale corrente, si calcola il rapporto di verosimiglianza con il quale si aggiornano le due statistiche locali. In corrispondenza dei flussi di dati non osservati all'istante temporale corrente, l'aggiornamento avviene rispettivamente aggiungendo e moltiplicando la costante 1 per le statistiche locali calcolate al tempo precedente. Aggiornando le due statistiche locali in questo modo si suppone che, in corrispondenza dei flussi di dati non osservati, il processo sia in controllo. Qualora lo schema non segnali un allarme, saranno selezionati i flussi di dati da osservare al tempo successivo applicando l'algoritmo TS. In particolare, se ne applica una versione semplificata che non valuta tutti i possibili sottoinsiemi di  $m$  flussi di dati, ottenibili a partire dai  $p$  flussi disponibili.

Una differente metodologia che impiega un approccio bayesiano per sorvegliare sequenzialmente il processo in presenza di osservazioni parziali e applica l'algoritmo TS viene proposta in Guo et al. (2022). I flussi di dati HD, acquisiti in tempo reale, vengono decomposti ed espressi sotto forma di espansione in funzioni di base. In particolare si impiegano due insiemi di funzioni di base e, di conseguenza, due vettori di coefficienti associati ad essi per descrivere, rispettivamente, l'andamento del processo in controllo e gli eventuali andamenti indesiderati che potrebbero verificarsi. La decomposizione è chiamata BSSCD e prevede l'impiego della distribuzione a priori *Spike-Slab* (Mitchell &

Beauchamp, 1988) per il vettore di coefficienti che descrive i possibili andamenti indesiderati. Le distribuzioni a posteriori dei vettori di coefficienti vengono stimate tramite un metodo di inferenza bayesiana variazionale che risulta più veloce dei metodi *Markov Chain Monte Carlo* (MCMC), pur ottenendo un'accuratezza comparabile (Wang & Blei, 2019). Impiegando le distribuzioni a posteriori stimate, si calcola il fattore di Bayes a posteriori (PBF, Aitkin, 1991) che rappresenta la statistica di controllo dello schema proposto. Quando il PBF eccede il limite di controllo lo schema segnala un allarme; al contrario, quando il processo viene dichiarato in controllo si impiega l'algoritmo TS per selezionare il sottoinsieme di flussi di dati da osservare al tempo successivo. Questa seconda metodologia combina, dunque, la decomposizione BSSCD con l'algoritmo TS (TS-BSSCD). L'algoritmo TS seleziona, ad ogni istante temporale, il sottoinsieme di flussi di dati che massimizza la statistica di controllo PBF, al fine di identificare, il più presto possibile, un'eventuale deviazione dallo stato in controllo del processo. La massimizzazione del PBF corrisponde, infatti, alla massimizzazione della funzione di vincita nel problema MAB. L'algoritmo TS è applicato sia al caso in cui il PBF sia non lineare, sia a quello semplificato in cui il PBF sia lineare, con conseguente riduzione del costo computazionale. Sfortunatamente, TS-BSSCD e TSSRP si basano sull'assunzione di indipendenza ed identica distribuzione dei flussi di dati; in aggiunta, TS-BSSCD richiede che i flussi di dati abbiano distribuzione normale. Come riportato in precedenza queste assunzioni possono rivelarsi fortemente irrealistiche nella maggior parte delle applicazioni reali. Inoltre, per applicare TSSRP è necessario specificare a priori le distribuzioni del processo IC e OC, e la metodologia potrebbe risultare non robusta in presenza di un'errata specificazione delle distribuzioni. La specificazione della distribuzione IC può risultare non banale nella pratica e ulteriori difficoltà si riscontrano nella specificazione della distribuzione OC assumendo, come in TSSRP, che sia possibile descrivere tutti i possibili andamenti fuori controllo che il processo può sperimentare impiegando un'unica distribuzione.

Allo scopo di sorvegliare in tempo reale flussi di dati eterogenei in distribuzione, in Ye et al. (2023) si propone una metodologia di sorveglianza non parametrica basata sull'anti-rango che impiega l'algoritmo TS come strategia di campionamento adattiva (AiTS). La metodologia non richiede che sia soddisfatta l'assunzione di normalità, né

di indipendenza ed identica distribuzione dei flussi di dati, né l'assunzione di scambiabilità. Tuttavia, la procedura proposta è stata sviluppata per sorvegliare flussi di dati indipendenti nel tempo e, pertanto, non è adeguata in presenza di flussi di dati autocorrelati. Lo schema di controllo AiTS impiega e generalizza la procedura CUSUM basata sull'anti-rango, proposta in Qiu & Hawkins (2001), per un processo generatore pienamente osservabile. Tale procedura valuta lo stato del processo sorvegliando il vettore che indica quale sia la più piccola misurazione (primo anti-rango) o, al contrario, la più grande misurazione (ultimo anti-rango) acquisita al tempo corrente. Tuttavia, in presenza di osservazioni parziali non è possibile determinare quale sia la più piccola o la più grande misurazione acquisita, poiché si osservano solo  $m$  misurazioni tra le  $p$  disponibili. Si consideri che l'interesse sia rivolto all'identificazione del flusso di dati in corrispondenza del quale si osserva la più grande misurazione acquisita all'istante temporale corrente. Assumendo di poter osservare solo  $m$  flussi di dati, con la procedura AiTS, in corrispondenza di ciascun flusso, si stima la probabilità che all'istante corrente il flusso di dati considerato restituisca la più grande misurazione tra le  $p$ . Anche in corrispondenza dei flussi di dati non osservati all'istante corrente, tali probabilità sono stimate per avere un'indicazione su quali siano i flussi di dati che mostrano un andamento sospetto. Il vettore delle probabilità stimate viene impiegato nella procedura CUSUM di Qiu & Hawkins (2001), al posto dell'indicatore dell'ultimo anti-rango e lo schema segnala un allarme se la statistica di controllo eccede il limite. Quando lo schema non segnala un allarme, l'algoritmo TS seleziona i flussi di dati da osservare al tempo successivo, in particolare, AiTS applica la versione semplificata dell'algoritmo che non valuta tutti i possibili sottoinsiemi di  $m$  flussi di dati, ottenibili a partire dai  $p$  totali.

La strategia di campionamento adattiva TS, quando impiegata nelle metodologie di sorveglianza sequenziale, rappresenta quindi un'efficace alternativa alle altre strategie impiegate per selezionare sequenzialmente i flussi di dati da osservare ad ogni istante temporale. Con la selezione effettuata dall'algoritmo, i principi di esplorazione e sfruttamento vengono bilanciati, migliorando notevolmente la reattività dello schema di controllo nel segnalare tempestivamente l'allarme (Guo et al., 2022). L'algoritmo è adattivo poiché le distribuzioni a posteriori, ciascuna relativa ad ogni *data stream*, vengono aggiornate sulla base delle nuove evidenze osservate. Inoltre, l'algoritmo TS costituisce

una valida alternativa alle altre strategie di campionamento adattive proposte in letteratura per risolvere il problema MAB, come ad esempio l'UCB (Chapelle & Li, 2011). Si rimanda a Lattimore & Szepesvári (2020) per dettagli e approfondimenti sull'algoritmo UCB; inoltre, si veda Zhang & Hoi (2019) per un'applicazione dell'algoritmo UCB al contesto della sorveglianza sequenziale in presenza di osservazioni parziali, in cui i flussi di dati non sono ad elevata dimensionalità.

L'algoritmo TS richiede la specificazione della distribuzione a priori associata a ciascun flusso di dati. La distribuzione a priori ha un impatto significativo sulla performance dell'algoritmo (Lattimore & Szepesvári, 2020). Nell'inferenza bayesiana classica, in presenza di una specificazione della distribuzione a priori non adeguata, la verosimiglianza avrà, a differenza della a priori, un notevole impatto sulla distribuzione a posteriori. Ciò non avviene nel problema MAB in cui, in presenza di una distribuzione a priori che sottostima fortemente la probabilità di successo associata ad una specifica leva, l'algoritmo TS potrebbe non selezionare mai tale leva e, dunque, non osservare nessuna realizzazione da essa (Lattimore & Szepesvári, 2020). Nel contesto della sorveglianza, la misspecificazione della distribuzione a priori potrebbe, quindi, compromettere fortemente la reattività dello schema di controllo. L'algoritmo TS potrebbe cioè non selezionare mai alcuni *data stream*, non esplorando a sufficienza nonostante sia necessario allocare sempre un numero sufficiente di risorse per l'esplorazione quando il processo generatore è non stazionario.

Quando la distribuzione a priori è coniugata per la verosimiglianza, l'aggiornamento della distribuzione a posteriori avviene in maniera efficace dal punto di vista computazionale, aggiornando esclusivamente i parametri della a priori, come nel caso del *Bernoulli* MAB. Tuttavia, nelle applicazioni reali, spesso si impiegano modelli più complessi per i quali l'inferenza bayesiana esatta è computazionalmente impraticabile (Russo et al., 2018). Pertanto, quando i flussi di dati sono caratterizzati da una forte struttura di correlazione spazio-temporale bisogna estendere l'algoritmo TS e impiegare algoritmi - ad esempio, di tipo MCMC - per campionare dalla distribuzione a posteriori approssimata (Zhang & Mei, 2022). Si rimanda a Russo et al. (2018) in cui si presentano alcuni approcci per generalizzare l'algoritmo TS campionando da un'approssimazione della distribuzione a posteriori. Sfortunatamente, non è stato ancora proposto in letteratura

un approccio bayesiano per la sorveglianza sequenziale di flussi di dati HD in presenza di osservazioni parziali che sorveglia i dati nel loro formato originale e campiona, ad ogni istante temporale e in maniera computazionalmente efficiente, da un'approssimazione della distribuzione a posteriori. Data la particolare struttura di correlazione spazio-temporale che caratterizza i flussi di dati, l'elevata dimensionalità e l'elevata frequenza di acquisizione dei *frame*, l'analisi e la sorveglianza in tempo reale del processo, di per sé dinamico, dovrebbero essere condotte sfruttando tutta l'informazione, spaziale e temporale, disponibile. La sorveglianza, l'imputazione delle misurazioni mancanti ed il campionamento potrebbero essere condotti impiegando i flussi di dati nel loro formato originale, senza richiedere che siano soddisfatti particolari assunti distributivi, difficili da soddisfare nella maggior parte delle applicazioni reali.



## Capitolo 3

# Casi reali di applicazione ed approccio proposto

Nell'ambito della sorveglianza condotta tramite video *frame*, saranno presentati, nel seguito, due casi reali: la sorveglianza in linea di un processo di produzione additiva del metallo e la sorveglianza in tempo reale dell'attività solare. I due casi reali sono riferiti a due contesti applicativi estremamente distanti tra loro, in quanto, il primo contesto si riferisce all'ambito delle tecnologie di fabbricazione additiva, caratterizzate da elevati standard di qualità, mentre il secondo riguarda l'osservazione di fenomeni fisici che possono causare danni alle infrastrutture tecnologiche.

Successivamente, si propone un approccio non parametrico finalizzato alla ricostruzione e al campionamento sequenziale di flussi di dati da video o sequenze di immagini allo scopo di sorvegliare il processo o, equivalentemente il fenomeno, in presenza di osservazioni parziali. L'approccio impiega la vicinanza spaziale dei flussi di dati per imputare le misurazioni mancanti e per assegnare, in corrispondenza di ciascun flusso di dati, ossia ciascun pixel, una probabilità di entrare a far parte del sottoinsieme di pixel che, in assenza di allarme, sarà osservato al tempo successivo.

## 3.1 Due casi reali

Di seguito si descrivono due casi reali riferiti, rispettivamente, alla sorveglianza del processo di produzione additiva *Selective Laser Melting* (SLM) e alla sorveglianza dell'attività solare. In particolare, nel primo caso, il processo SLM viene sorvegliato tramite video *frame*, ciascuno di dimensione  $121 \times 71$  pixel, acquisiti ad elevata frequenza per identificare l'occorrenza di gravi anomalie termiche locali, chiamate *hotspot*.

Nel secondo caso l'attività solare viene sorvegliata tramite una sequenza di 300 immagini, ciascuna di dimensione  $292 \times 232$  pixel, acquisite allo scopo di identificare i brillamenti solari che possono verificarsi sulla superficie del Sole.

### 3.1.1 *Hotspot* nei processi di produzione additiva del metallo

I processi di produzione additiva del metallo permettono la realizzazione di prodotti con forme innovative, dettagli complessi e strutture leggere, che risulterebbero difficili da realizzare utilizzando i tradizionali sistemi di produzione (Colosimo & Grasso, 2018). Il *Selective Laser Melting* (SLM) è una tecnologia di AM appartenente alla categoria dei processi di fusione su letto di polvere (*Laser Powder Bed Fusion*, L-PBF) in cui un raggio laser fonde selettivamente uno strato di polvere metallica producendo, strato dopo strato, gli oggetti desiderati. La produzione strato dopo strato che caratterizza i processi di AM permette di acquisire e misurare in tempo reale indicatori della qualità del prodotto e della stabilità del processo, durante la fase di produzione (Yan et al., 2022). I sistemi di visione artificiale (MVS) consentono di acquisire immagini e video durante la produzione di ogni strato e di misurare, dunque, in tempo reale le cosiddette "firme" (*signature*) del processo. Il termine "firma" del processo è stato utilizzato per la prima volta in Mani et al. (2015) e si riferisce ad una o più grandezze che possono essere misurate durante la fase di produzione per raccogliere informazioni riguardanti la stabilità del processo, i fenomeni fisici sottostanti e l'insorgenza di possibili difetti. Durante la scansione con il raggio laser, è possibile acquisire immagini in tempo reale ed osservare rapidi fenomeni transitori legati alla condizione termica del processo ed emissioni di scintille e plasma (Grasso et al., 2021). Ad esempio, la diffusione delle



scintille rappresenta una firma del processo e fornisce informazioni sullo stato del processo e sull'occorrenza di eventuali difetti volumetrici (porosità) all'interno del prodotto (Colosimo et al., 2022).

Con l'obiettivo di valutare in tempo reale la condizione termica del processo, diversi approcci sono stati sviluppati per sorvegliare il processo in linea, tramite immagini acquisite ad elevata frequenza, allo scopo di identificare gli *hotspot*; si veda per esempio Yan et al. (2022); Yang et al. (2022); Bugatti & Colosimo (2021); Colosimo & Grasso (2018); Grasso et al. (2017). Gli *hotspot* sono gravi anomalie di surriscaldamento locali che possono verificarsi durante la produzione di ciascuno strato del prodotto. In particolare, una regione affetta da un *hotspot* rimane calda per un periodo di tempo più lungo ed è caratterizzata da un tasso di raffreddamento più lento rispetto alle zone di fusione regolare, inducendo così delle deformazioni geometriche locali nei prodotti (Colosimo & Grasso, 2018). L'identificazione e la localizzazione degli *hotspot* consente di interrompere tempestivamente il processo per evitare di realizzare parti difettose e consente, inoltre, di ridurre il numero di difetti o, addirittura, riparare i difetti dei prodotti durante la fase di produzione.

I dati descritti in seguito sono stati impiegati in Colosimo & Grasso (2018) e si riferiscono al processo di produzione di tre porzioni triangolari, appartenenti alla forma complessa in Figura 3.1 (a). In particolare, le tre porzioni triangolari sono rappresentate in Figura 3.1 (b) e in essa si evidenziano gli angoli acuti in corrispondenza dei quali si possono osservare degli *hotspot*. Infatti, gli *hotspot* si presentano principalmente in corrispondenza di angoli acuti, sporgenze o pareti sottili della porzione realizzata, a causa di un accumulo eccessivo di calore che può verificarsi spesso in queste zone. In questo caso gli angoli presentano difetti e deformazioni locali, evidenziati nella Figura 3.1 (c). La Figura 3.1 è riportata in Colosimo & Grasso (2018).

Per sorvegliare il processo ed identificare gli *hotspot* vengono acquisite, ad elevata frequenza, immagini in scala di grigi di dimensione  $121 \times 71$  pixel, durante la produzione degli strati di ciascuna porzione triangolare. Nella Figura 3.2 viene rappresentata una sequenza di *frame* acquisita durante la produzione, tramite SLM, di uno strato di una porzione triangolare. Le sequenze di immagini catturano, in tempo reale, la fase di produzione di ciascuno strato del prodotto che è caratterizzata da un andamento

dinamico, anche quando il processo è in controllo. In particolare, quando il processo è in controllo i *frame* catturano il raggio laser, le zone di fusione regolare e le emissioni derivanti dall'interazione tra il raggio laser e la polvere metallica depositata. L'approssimazione e la stima della distribuzione del processo in controllo che descrive l'andamento dinamico del processo in condizioni di stabilità è non banale e la maggior parte delle metodologie tradizionali di sorveglianza sequenziale non risultano in questo caso adeguate.

La sorveglianza sequenziale del processo, in questo contesto, si traduce nell'analisi in tempo reale della luminosità associata a ciascun pixel che compone ciascuna immagine acquisita. In particolare, i pixel possono essere suddivisi in tre categorie: (i) pixel di *background*, in corrispondenza dei quali non avviene alcuna interazione tra il raggio laser e la polvere metallica, (ii) pixel corrispondenti alle zone di fusione regolare che rappresentano lo stato in controllo del processo e (iii) pixel corrispondenti agli *hotspot* ossia i fenomeni di surriscaldamento locale che possono verificarsi durante la fase di produzione. Un *hotspot* ha dimensioni ridotte ed è dunque rappresentato da un ridotto numero di pixel adiacenti tra loro, la cui luminosità è

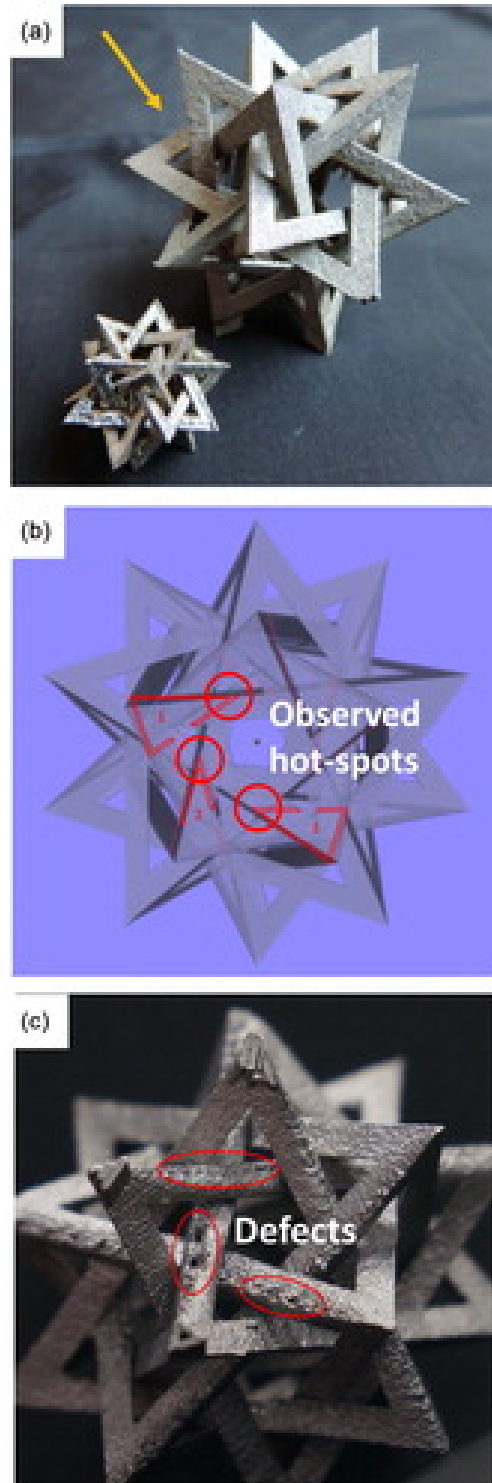


FIGURA 3.1: (a) forma complessa realizzata dal processo SLM; (b) esempi di porzioni triangolari; (c) difetti locali negli angoli acuti delle porzioni triangolari.

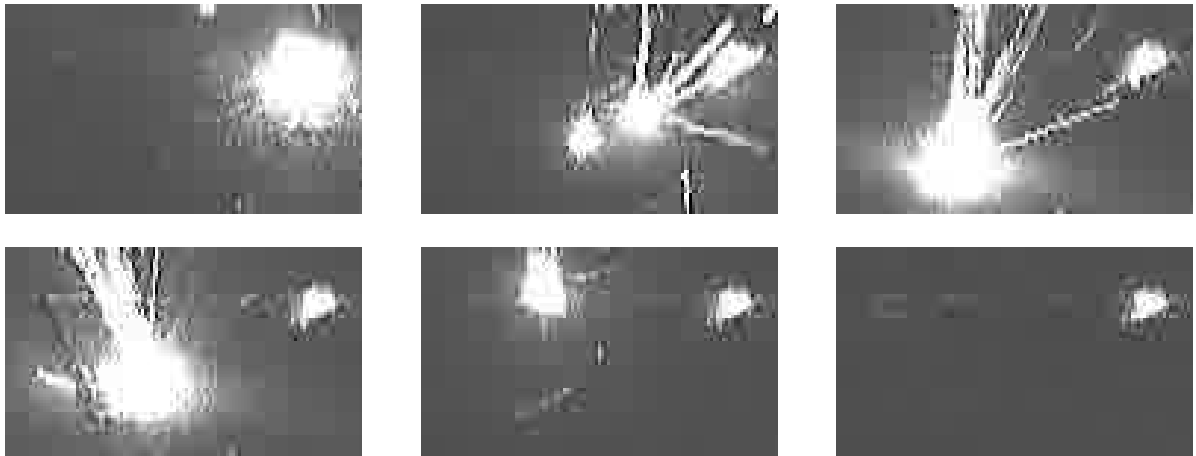


FIGURA 3.2: Sequenza di *frame* estratta da un video acquisito durante il processo SLM di una porzione triangolare.

intensa e costante per un lungo periodo di tempo. Valutando l'andamento del processo tramite video *frame*, il cambiamento indesiderato (*hotspot*) che il processo sperimenta è sparso, poiché influenza un ridotto numero di pixel, ed è persistente nel tempo, poiché la luminosità registrata in corrispondenza di questi pixel è intensa e rimane elevata per un periodo di tempo più lungo rispetto alle zone di fusione regolare.

Quando sono disponibili solo osservazioni parziali del processo ad ogni istante temporale è osservabile solo un ridotto numero di pixel ( $m < p$ ) da ogni *frame* acquisito. In tal caso, le tecniche di sorveglianza, oltre alla costruzione di uno schema efficace per la rilevazione degli *hotspot*, devono fronteggiare le ulteriori difficoltà derivanti dalla presenza dei dati mancanti e dalla necessità di impiegare una strategia di campionamento adattiva.

### 3.1.2 Il fenomeno dei brillamenti solari

I brillamenti solari sono definiti come improvvise, rapide ed intense variazioni della luminosità che si verificano sulla superficie del Sole (Augusto et al., 2011). I brillamenti solari, o eruzioni solari, emettono un gran numero di particelle cariche di energia che possono potenzialmente causare guasti alle reti elettriche o alle comunicazioni radio. Pertanto, individuare tempestivamente i brillamenti solari è estremamente rilevante al fine di intraprendere, in breve tempo, azioni preventive e correttive. Tuttavia, la temperatura della superficie del Sole varia gradualmente nel tempo e i brillamenti solari

avvengono sporadicamente ed interessano piccole regioni della superficie solare. Similmente agli *hotspot*, il fenomeno dei brillamenti solari che si verifica sulla superficie del Sole si manifesta come un cambiamento indesiderato sparso, poiché coinvolge un numero di pixel estremamente ridotto, persistente nel tempo e caratterizzato da una forma irregolare. Anche in questo caso, la sparsità nel tempo e nello spazio, che caratterizza il fenomeno dei brillamenti solari, rappresenta una sfida per le tecniche di sorveglianza sequenziale tradizionali (Yan et al., 2018).

Per sorvegliare l'attività solare e identificare i brillamenti solari, sono stati acquisiti e resi disponibili 300 *frame*, ciascuno di dimensione  $292 \times 232$  pixel, raccolti dal satellite *Solar Dynamics Observatory* (NASA, 2009). Osservando tutti i pixel che compongono ciascuna immagine della sequenza e, in particolare, con la metodologia ST-SSD (Yan et al., 2018) sono stati identificati tre brillamenti solari che avvengono agli istanti temporali [191, 194], [216, 237] e [257, 258] e sono catturati dai *frame* corrispondenti. In Figura 3.3 sono riportati i *frame* [216, 219] in scala di grigi. Nella sequenza riportata emerge un brillamento solare, il quale si manifesta attraverso un gruppo di pixel adiacenti tra loro che mostrano un livello di luminosità elevato, che persiste anche nei *frame* successivi.

Le immagini sono acquisite ad elevata risoluzione e, a causa dell'elevata dimensionalità che caratterizza ciascun *frame* della sequenza, le metodologie di sorveglianza tradizionali potrebbero essere inefficaci nell'identificare tempestivamente i brillamenti solari, a causa della limitata banda di trasmissione dei dati al centro di analisi e delle limitate risorse computazionali disponibili per l'analisi in tempo reale (Gómez et al., 2022). Di conseguenza, è possibile analizzare in tempo reale solo un ridotto numero di pixel  $m$  ad ogni istante temporale, ossia in ogni *frame* acquisito. In presenza di osservazioni parziali ad ogni istante temporale, oltre ad un efficace schema di controllo, si necessita dunque di un metodo di imputazione dei dati mancanti e di una strategia di campionamento adattiva che bilanci l'esplorazione e lo sfruttamento delle risorse, in un contesto di risorse limitate. Allo scopo di identificare tempestivamente i brillamenti solari in presenza di osservazioni parziali, le metodologie TRAS (Liu et al., 2015), R-SADA (Xian et al., 2021), SASAM (Wang et al., 2018), TSS (Gómez et al., 2022), TSSRP (Zhang & Mei, 2022) e TS-BSSCD (Guo et al., 2022), descritte nei Capitoli 1 e 2, sono

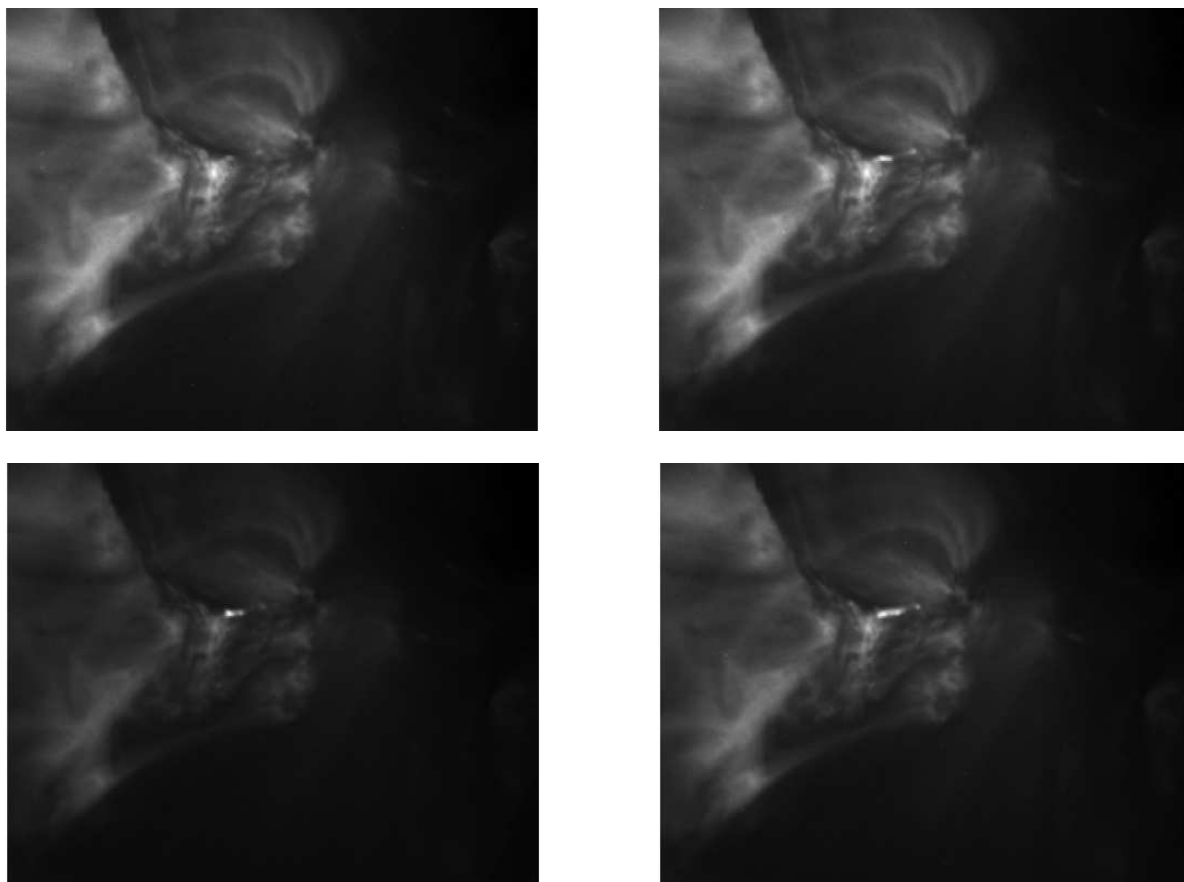


FIGURA 3.3: Sequenza di *frame* [216, 219] in scala di grigi contenuti nella sequenza originale.

state applicate alla sequenza di immagini acquisita dal satellite. Tali metodologie, ad eccezione della TSS, assumono che i flussi di dati siano indipendenti ed identicamente distribuiti ed impiegano l'algoritmo MOUSSE (*Multiscale Online Union of SubSpaces Estimation*, Xie et al., 2012) per ricavare in tempo reale i residui da sorvegliare, i quali sono approssimativamente indipendenti e normalmente distribuiti.

## 3.2 L'approccio proposto

Con l'obiettivo di sviluppare uno schema di controllo per la sorveglianza sequenziale di video *frame* parzialmente osservabili, impiegando un approccio non parametrico, si propone qui un metodo di imputazione delle misurazioni mancanti ed una strategia di campionamento per selezionare, ad ogni istante temporale e in assenza di allarme,  $m$  flussi di dati da osservare all'istante temporale successivo. In particolare, non si richiede che siano soddisfatte assunzioni distributive come l'assunzione di indipendenza

ed identica distribuzione o l'assunzione di normalità dei flussi di dati, ma si impiegano i flussi di dati acquisiti nel loro formato originale. Ciascun *frame* di dimensione  $M \times N$ , acquisito in ogni istante temporale, può essere considerato come una griglia in cui ogni cella corrisponde ad un pixel. Siano  $p$  i pixel che compongono ciascuna immagine acquisita, con  $p = MN$ . In presenza di osservazioni parziali, per ogni *frame* acquisito, si osservano solo  $m$  pixel dei  $p$  totali ( $m < p$ ) e la luminosità, osservata in corrispondenza di ciascun pixel, può essere considerata come la terza dimensione da associare ad ogni cella che compone la griglia. Al primo istante temporale, dal primo *frame* acquisito, si osservano  $m$  pixel selezionati in maniera casuale. Impiegando dei *frame* in scala di grigi, le misurazioni della luminosità variano nell'intervallo  $[0, 1]$ , dove 0 corrisponde al colore nero e 1 al colore bianco. L'approccio proposto prevede di identificare, in ogni istante temporale, i pixel in corrispondenza dei quali si osservano le misurazioni massime tra le  $m$  misurazioni acquisite, identificando dunque quelle che si suppone siano le mode della distribuzione sottostante parzialmente osservabile. Le mode identificate vengono successivamente impiegate nella fase di imputazione delle misurazioni mancanti e nella fase di campionamento.

### 3.2.1 Identificazione delle mode

Sia  $\mathbf{X}_t \in \mathbb{R}^p$ , il vettore  $p$ -dimensionale delle misurazioni acquisite al tempo  $t$  che, in presenza di osservazioni parziali, contiene  $m$  misurazioni della luminosità osservata in corrispondenza di  $m$  pixel e  $p - m$  misurazioni mancanti. Le misurazioni contenute in  $\mathbf{X}_t$  vengono ordinate in maniera decrescente, ottenendo  $X_{(1)}, \dots, X_{(p)}$  con  $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(p)}$ , dove  $X_{(m+1)}, \dots, X_{(p)}$  corrispondono alle misurazioni mancanti. Sia  $\mathbf{X}_t^o$  il vettore delle misurazioni ordinate e siano  $e_{(1)}, \dots, e_{(p)}$  le etichette dei pixel corrispondenti alle misurazioni  $X_{(1)}, \dots, X_{(p)}$ .

Allo scopo di identificare le mode della distribuzione sottostante, sulla base delle misurazioni osservate, si calcolano le differenze tra le misurazioni consecutive contenute nel vettore  $\mathbf{X}_t^o$ . In particolare, si calcolano le differenze tra misurazioni consecutive fin quando la differenza successiva è strettamente maggiore della differenza corrente. Ad esempio, qualora

$$X_{(1)} - X_{(2)} < X_{(2)} - X_{(3)} \quad (3.1)$$

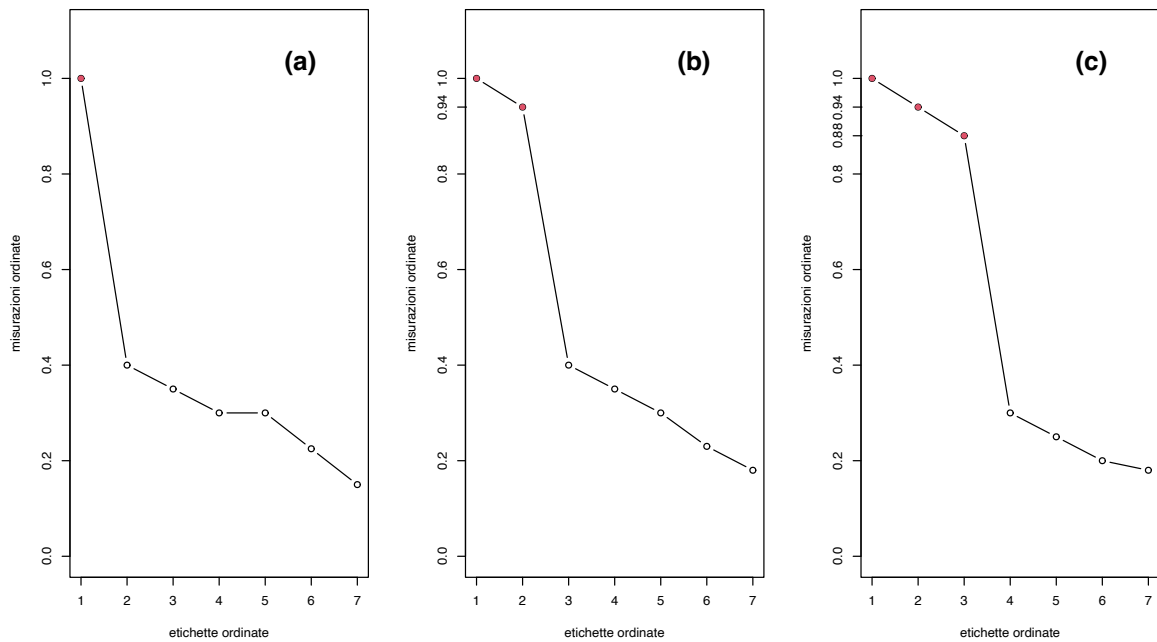


FIGURA 3.4: Esempio di identificazione di una moda (a), due mode (b) e tre mode (c) a partire da tre vettori di misurazioni osservate ed ordinate in maniera decrescente.

si identificano due mode in corrispondenza dei pixel aventi etichette  $e_{(1)}$  ed  $e_{(2)}$ , dai quali sono state acquisite rispettivamente le misurazioni  $X_{(1)}$  ed  $X_{(2)}$ .

La Figura 3.4 riporta un esempio di identificazione delle mode a partire da tre vettori di misurazioni osservate e ordinate. In particolare, la Figura 3.4 evidenzia i casi in cui si identificano, rispettivamente, una (a), due (b) e tre mode (c).

Il numero di mode non è fissato a priori ma viene calcolato automaticamente, confrontando le differenze tra misurazioni consecutive e la procedura di identificazione delle mode termina non appena la differenza calcolata è strettamente maggiore della differenza calcolata precedentemente. La procedura viene ripetuta in ogni istante temporale e l'identificazione delle mode ha impatto sull'imputazione delle misurazioni mancanti e sull'assegnazione, in corrispondenza di ciascun pixel, della probabilità di essere campionato all'istante temporale successivo.

### 3.2.2 Metodo di imputazione

Una volta identificate le mode, si calcola la distanza euclidea tra ciascun pixel, ossia tra ciascuna cella che compone la griglia, e la moda più vicina. La distanza euclidea tra un

pixel e la moda più vicina ad esso è pari a 0 quando, in corrispondenza di tale pixel, è stata identificata una moda all'istante temporale corrente. Sulla base delle distanze calcolate e, dunque, sfruttando la vicinanza spaziale dei pixel, si imputano le  $p - m$  misurazioni mancanti. In particolare, è ragionevole supporre che in corrispondenza dei pixel non osservati all'istante temporale corrente, ma adiacenti alle mode, la luminosità sia leggermente inferiore rispetto alla luminosità osservata in corrispondenza della moda più vicina. Pertanto, in ogni istante temporale, si imputano le misurazioni mancanti  $X_{(m+1)}, \dots, X_{(p)}$  supponendo che la luminosità, in corrispondenza di ciascun pixel non osservato all'istante corrente, diminuisca all'aumentare della distanza dalla moda più vicina. Per imputare le misurazioni mancanti, si impiega una funzione nucleo (*kernel*)  $K(\cdot)$ , ossia una funzione di densità di probabilità, tipicamente simmetrica, per assegnare un peso in corrispondenza di ogni pixel non osservato, sulla base della sua distanza dalla moda più vicina. In particolare, si consideri un pixel non osservato all'istante corrente e sia  $d$  la distanza euclidea del pixel considerato dalla moda più vicina ad esso. Si impiega la funzione *kernel* di Epanechnikov (Epanechnikov, 1969), di seguito denominata semplicemente come il *kernel* di Epanechnikov  $K_E(\cdot)$ , definita come

$$K_E(d, h_1) = \left(\frac{3}{4h_1}\right) \left(1 - \left(\frac{d}{h_1}\right)^2\right) \quad \text{per} \quad \left|\frac{d}{h_1}\right| \leq 1 \quad (3.2)$$

dove  $h_1 > 0$  è l'ampiezza di banda. In alternativa al *kernel* di Epanechnikov, è possibile impiegare altre funzioni *kernel* per assegnare un peso ai pixel non osservati come, ad esempio, il *kernel* Gaussiano. A differenza di quest'ultimo, il *kernel* di Epanechnikov ha supporto compatto ossia è diverso da 0 nell'intervallo finito  $(-h_1, h_1)$  centrato nella moda (Hastie et al., 2009). Pertanto, impiegando  $K_E(\cdot)$  si assegna un peso pari a 0 in corrispondenza dei pixel non osservati che si trovano ad una distanza dalla moda  $d$  maggiore o uguale ad  $h_1$  e, viceversa, si assegna un peso maggiore di 0 quando  $d$  è minore di  $h_1$ . In questo modo, si imputano le misurazioni della luminosità assegnando un valore pari a 0, ossia assegnando il colore nero, in corrispondenza dei pixel che hanno una distanza maggiore o uguale ad  $h_1$  dalla moda più vicina. Viceversa, valori compresi tra 0 ed il valore della luminosità della moda più vicina, sono assegnati in corrispondenza dei pixel aventi distanza dalla moda minore di  $h_1$ .



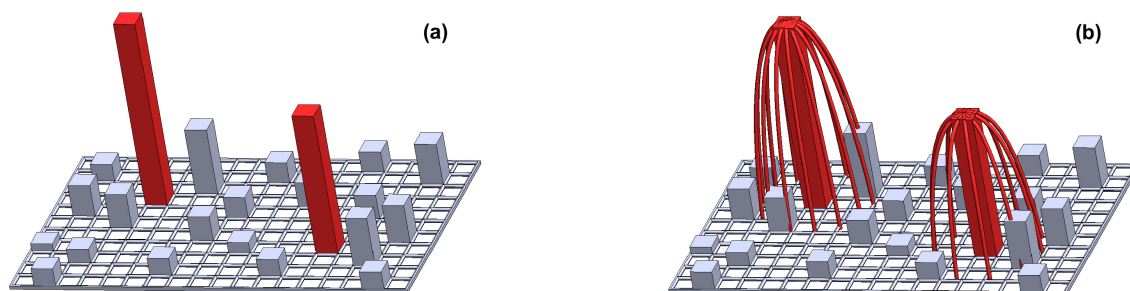


FIGURA 3.5: Esempio di identificazione di due mode (a), indicate dai parallelepipedi in rosso, e di imputazione delle misurazioni mancanti (b).

La Figura 3.5 riporta un esempio di identificazione di due mode (a) e di imputazione delle misurazioni mancanti (b) con il metodo proposto.

L'ampiezza di banda determina l'ampiezza della regione, costruita attorno a ciascuna moda identificata, all'interno della quale il *kernel* assegna peso non nullo. All'aumentare dell'ampiezza di banda, la regione è più ampia, pertanto, un numero maggiore di pixel riceve peso non nullo (Altman, 1992). Si noti che, una volta imputate le misurazioni mancanti, le misurazioni della luminosità in corrispondenza di tutti i pixel che compongono il *frame* non integrano a 1. Non integrano a 1 nemmeno le misurazioni della luminosità in corrispondenza dei pixel che hanno distanza minore di  $h_1$  dalla moda più vicina. Infatti, non si richiede che le misurazioni della luminosità soddisfino le proprietà di una densità di probabilità. Pertanto, nell'imputazione delle misurazioni mancanti, si assegnano luminosità che decrescono all'aumentare della distanza dalla moda senza imporre particolari vincoli, come l'integrazione a 1 delle misurazioni. Nel caso presente, all'aumentare dell'ampiezza di banda  $h_1$ , un numero maggiore di pixel riceve peso non nullo ma la luminosità attribuita a ciascun pixel non diminuisce all'aumentare di  $h_1$ ; diversamente avviene quando si impiegano gli stimatori *kernel* per stimare, in maniera non parametrica, una funzione di regressione o di densità.

Il valore dell'ampiezza di banda  $h_1$  deve essere fissato a priori. Questo valore deve essere selezionato misurando le dimensioni (in pixel) di una regione luminosa che possiede una grandezza media rispetto alle regioni luminose osservate nella maggior parte dei *frame* disponibili, acquisiti in precedenza. Per misurare le dimensioni della regione luminosa, è possibile inscrivere in un rettangolo, misurando altezza e larghezza

di quest'ultimo e selezionando  $h_1$  pari alla metà della dimensione inferiore tra le due. Dagli studi di simulazione condotti, selezionando il valore dell'ampiezza di banda in questo modo ed applicando il metodo di imputazione ad un'intera sequenza di *frame*, con  $h_1$  fissato, le ricostruzioni ottenute sono soddisfacentemente vicine ai *frame* originali. L'imputazione delle misurazioni mancanti e, dunque, il valore dell'ampiezza di banda, selezionato a priori, non ha impatto sulla successiva fase di campionamento che viene condotta supponendo che non sia stato segnalato un allarme.

### 3.2.3 Strategia di campionamento

In assenza di allarme, in corrispondenza di ciascun pixel tra i  $p$  disponibili, si assegna una probabilità di entrare a far parte del campione di  $m$  pixel che sarà osservato all'istante temporale successivo. Sia  $\mathbf{W}_t \in \mathbb{R}^p$  il vettore  $p$ -dimensionale delle probabilità assegnate, al tempo  $t$  in corrispondenza di ciascun pixel, di entrare a far parte del campione osservato al tempo  $t + 1$ . Per assegnare le probabilità e, dunque, determinare tutti gli elementi del vettore  $\mathbf{W}_t$  si impiegano le mode identificate al tempo  $t$  con la procedura descritta al paragrafo 3.2.1. Si impiega poi un *kernel* a code pesanti, in particolare un *kernel*  $t$  di Student con ampiezza di banda  $h_2 > 0$ , centrato in ciascuna moda identificata. Si consideri un pixel tra i  $p$  disponibili e sia  $d$  la distanza del pixel considerato dalla moda più vicina ad esso. Sia  $n$  il numero di gradi di libertà, il *kernel*  $t$  di Student  $K_t(\cdot)$  è definito come

$$K_t(d, n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{d^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} \quad \text{con} \quad n = \frac{2h_2^2}{h_2^2 - 1} \quad (3.3)$$

dove  $\Gamma(\cdot)$  è la funzione gamma. Impiegando il *kernel*  $K_t(\cdot)$ , sulla base della distanza di ciascun pixel dalla moda più vicina, si determinano gli elementi del vettore  $\mathbf{W}_t$  in modo che la loro somma sia pari a 1. Le probabilità assegnate decrescono all'aumentare della distanza dalla moda più vicina, pertanto, i pixel che si trovano nelle vicinanze delle mode identificate all'istante corrente saranno selezionati all'istante successivo con probabilità maggiore rispetto ai pixel distanti dalle mode. Al tempo  $t + 1$ , con probabilità elevata, saranno acquisite le misurazioni della luminosità nelle regioni adiacenti alle mode, favorendo il principio dello sfruttamento delle risorse e allo stesso tempo,

impiegando un *kernel* con code pesanti, saranno acquisite anche alcune misurazioni in corrispondenza di pixel distanti dalle mode, favorendo l'esplorazione delle risorse.

Il valore dell'ampiezza di banda  $h_2$  deve essere selezionato a priori, scegliendo un valore elevato, approssimativamente pari al massimo tra la metà del numero di righe e la metà del numero di colonne dei *frame* acquisiti. Scegliendo un valore elevato per l'ampiezza di banda  $h_2$ , si assegna una probabilità non nulla e non trascurabile anche in corrispondenza dei pixel piuttosto distanti dalle mode identificate all'istante corrente, continuando ad esplorare anche nelle regioni spaziali distanti dalle mode. Le probabilità assegnate al tempo  $t$ , contenute nel vettore  $\mathbf{W}_t$ , sommano a 1 e all'aumentare del valore dell'ampiezza di banda  $h_2$ , le probabilità assegnate diminuiscono in maniera proporzionale. Tuttavia, si consiglia di selezionare il valore di  $h_2$  scegliendo un valore elevato poiché è cruciale continuare ad esplorare, al fine di rilevare ulteriori regioni spaziali sospette, che mostrano un livello di luminosità elevato, indice di un possibile andamento fuori controllo. Allo stesso tempo, la strategia di campionamento seleziona i pixel adiacenti alle mode; queste vengono identificate in ogni istante temporale, senza impiegare l'informazione relativa alle mode identificate in precedenza, in modo da non soffermarsi nelle regioni adiacenti alle mode identificate inizialmente. La strategia di campionamento è in grado di selezionare pixel appartenenti a regioni multiple dello spazio, in cui l'andamento del processo diviene sospetto e indesiderato, ed è in grado di campionare simultaneamente in ognuna di esse. In presenza di regioni dalle dimensioni ridotte e dalla luminosità elevata e persistente nel tempo, *frame* dopo *frame*, i pixel campionati dalla strategia si concentrano nelle zone contenenti le regioni luminose, talvolta osservando queste ultime interamente.

Nell'implementazione dell'approccio proposto ai due casi reali, riportata nel Capitolo 4, si mostrano i risultati ottenuti impiegando dei valori di  $h_1$  ed  $h_2$ , selezionati a priori in ciascuno dei due casi reali e che vengono mantenuti costanti per tutti i *frame* appartenenti a ciascuna delle due sequenze disponibili.



# Capitolo 4

## Applicazione dell'approccio proposto

L'approccio proposto, illustrato nel Capitolo 3, viene applicato ai due casi reali relativi all'identificazione degli *hotspot* e dei brillamenti solari. Supponendo che solo un ridotto numero di pixel sia osservabile in ogni istante temporale, si applica l'approccio proposto a diversi *frame* e a sequenze di questi, estratti dai tre video disponibili e dalla sequenza originale raccolta dal satellite. L'approccio proposto è stato implementato utilizzando il *software* R (R Core Team, 2022), con il codice riportato in Appendice.

### 4.1 Identificazione degli *hotspot*

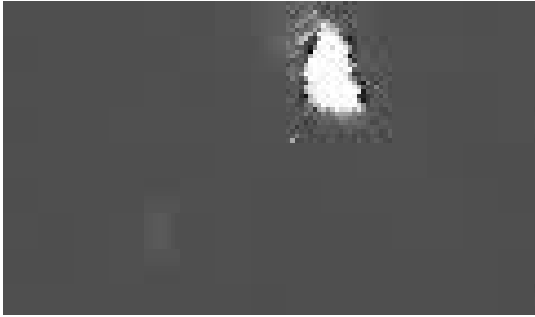
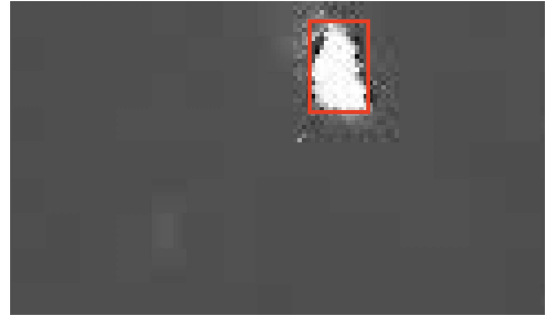
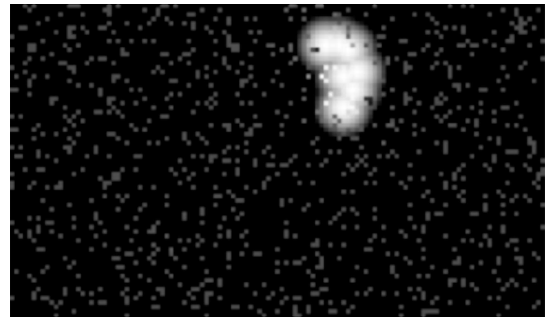
I *frame* riportati in seguito sono stati estratti, uno al secondo, dai video impiegati in Colosimo & Grasso (2018). Ciascun *frame* ha dimensione  $121 \times 71$  pixel e la risoluzione non è particolarmente elevata. Assumendo di poter osservare in ogni istante temporale solo  $m = 1000$  pixel tra i  $p$  disponibili, con  $p = 8591$ , si imputano le misurazioni mancanti applicando l'approccio proposto prima ad un *frame* che contiene un'unica regione luminosa, dalla forma piuttosto regolare, fornendo indicazioni sulla scelta dell'ampiezza di banda  $h_1$ . Successivamente, selezionando  $m$  pixel in maniera casuale, il metodo di imputazione viene applicato a *frame* contenenti scintille, più di una regione luminosa ed un'unica regione luminosa di dimensioni ridotte.

Si riporta, inoltre, la selezione dei pixel effettuata dalla strategia di campionamento proposta, applicata ai *frame* contenenti più di una regione luminosa e la sola regione luminosa di dimensioni ridotte. Il metodo di imputazione delle misurazioni mancanti e la strategia di campionamento sono infine applicati ad una sequenza di due *frame* consecutivi, estratti dal primo dei tre video disponibili.

### 4.1.1 Scelta dell'ampiezza di banda

Per selezionare il valore dell'ampiezza di banda  $h_1$ , impiegata nella fase di imputazione delle misurazioni mancanti, si impiega il *frame* rappresentato nella Figura 4.1. Il *frame* ritrae una regione luminosa, corrispondente ad un angolo acuto della porzione triangolare in cui si verifica un *hotspot*. La regione luminosa è di medie dimensioni e la sua forma è piuttosto regolare, rispetto alle altre regioni nei *frame* disponibili. Pertanto, impiegando questo *frame* è agevole selezionare un valore per l'ampiezza di banda  $h_1$ , che sarà poi impiegato in tutti i *frame* estratti dai tre video acquisiti durante il processo SLM. L'ampiezza di banda  $h_1$  deve essere selezionata scegliendo un valore approssimativamente pari alla metà della dimensione inferiore tra le due dimensioni, altezza e larghezza, misurate in pixel, della regione luminosa rappresentata nella Figura 4.1. Per identificare le dimensioni della regione luminosa e selezionare il valore dell'ampiezza di banda  $h_1$ , è possibile tracciare il più piccolo rettangolo (*bounding box*) che circonda la regione, rappresentato in rosso nella Figura 4.2, e misurarne le dimensioni. In questo caso, la dimensione inferiore tra le due è la larghezza e il valore selezionato per l'ampiezza di banda  $h_1$  è pari a 7 che coincide con la metà della larghezza del rettangolo. Gli studi effettuati selezionando un valore di  $h_1$  compreso tra 5 e 9 portano ad una ricostruzione del *frame* originale ancora soddisfacente (Figura 4.5). Alla luce dei risultati ottenuti, il valore di  $h_1 = 7$  sarà impiegato per tutte le ricostruzioni riportate nei paragrafi successivi, relativi all'applicazione dell'approccio proposto per l'identificazione degli *hotspot* nel processo SLM.

Supponendo di essere al primo istante temporale della sorveglianza, si campionano in maniera casuale  $m = 1000$  pixel tra  $p$  disponibili (Figura 4.3). Con l'obiettivo di ricostruire l'intero *frame* a partire dalle  $m$  misurazioni acquisite, si applica l'approccio proposto nel Capitolo 3, procedendo con l'identificazione delle mode e l'imputazione delle

FIGURA 4.1: Video 2 - *frame* 5.FIGURA 4.2: Video 2 - *frame* 5 e *bounding box* in rosso.FIGURA 4.3: Campionamento casuale di  $m = 1000$  pixel.FIGURA 4.4: Ricostruzione del *frame* sulla base delle  $m$  misurazioni acquisite ( $h_1 = 7$ ).

$p-m$  misurazioni mancanti, sulla base della distanza di ciascun pixel non osservato dalla moda più vicina. Impiegando il *kernel* di Epanechnikov ed il valore  $h_1 = 7$ , si ricostruisce il *frame* originale, ottenendo la ricostruzione riportata nella Figura 4.4. L'imputazione delle misurazioni mancanti avviene assegnando un valore della luminosità pari a 0, ossia assegnando il colore nero, in corrispondenza dei pixel che si trovano ad una distanza maggiore o uguale ad  $h_1$  dalla moda più vicina. Si ricordi che l'imputazione delle misurazioni mancanti e, dunque, il valore selezionato per l'ampiezza di banda  $h_1$  non ha impatto sulla fase di campionamento, la quale impiega le mode identificate e non le misurazioni imputate.

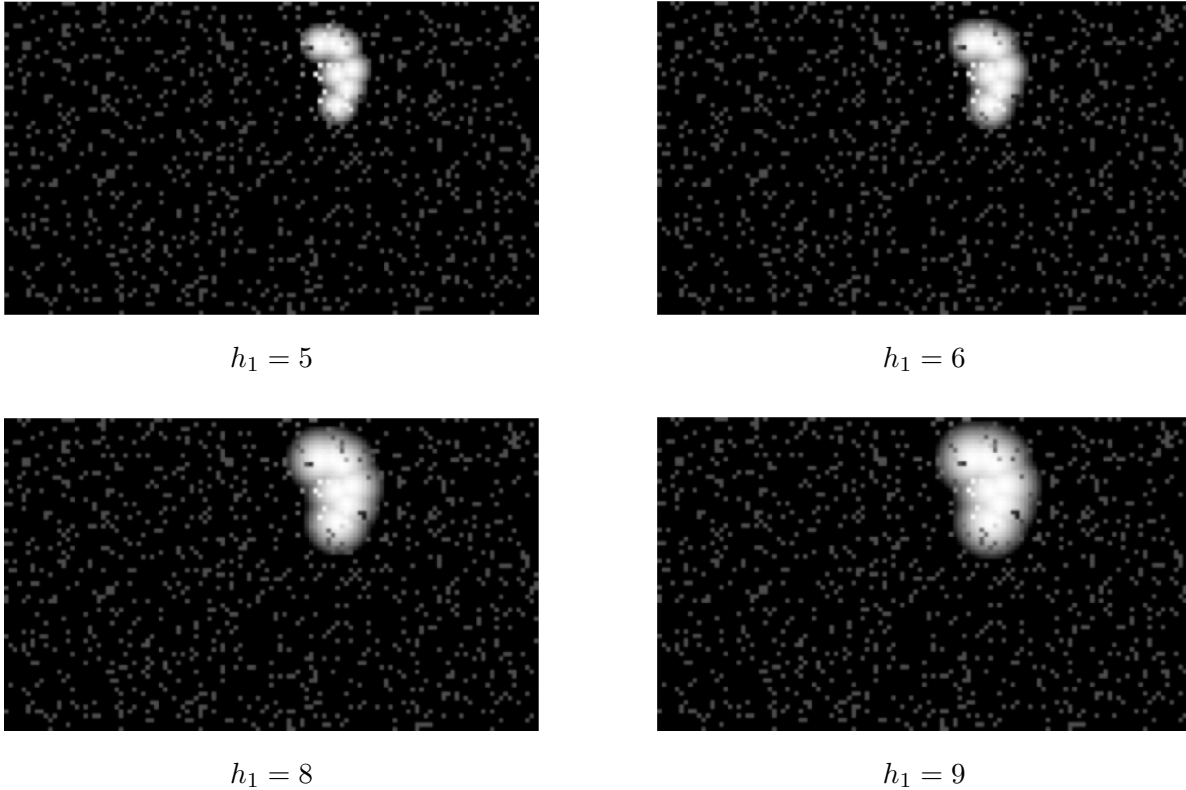


FIGURA 4.5: Ricostruzioni del *frame* impiegando valori di  $h_1$  diversi da 7, appartenenti all'intervallo consigliato.

### 4.1.2 Scintille e loro riconoscimento

Le sequenze di immagini estratte dai video acquisiti durante il processo SLM oltre che catturare l'andamento dinamico del processo, catturano anche la diffusione delle scintille generate dall'interazione tra il raggio laser e la polvere metallica depositata. L'imputazione delle misurazioni mancanti si basa sulle mode identificate e sulla distanza di ciascun pixel non osservato dalla moda più vicina. Qualora si identifichino una o più mode in corrispondenza di pixel aventi luminosità elevata ma relativi alle scintille, le misurazioni mancanti verranno imputate adottando lo stesso criterio impiegato per imputare le regioni luminose, relative alle zone di fusione regolare e agli *hotspot*.

Ad esempio, il *frame* nella Figura 4.6 cattura una zona di fusione regolare e le scintille generate dall'interazione tra il raggio laser e la polvere metallica depositata. Osservando solo  $m = 1000$  pixel, campionati in maniera casuale dal *frame* originale ed impiegando  $h_1 = 7$ , si imputano le misurazioni mancanti sulla base delle mode identificate, ottenendo la ricostruzione riportata nella Figura 4.7. Imputando le misurazioni

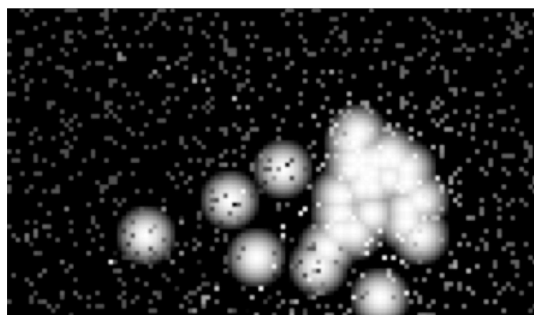
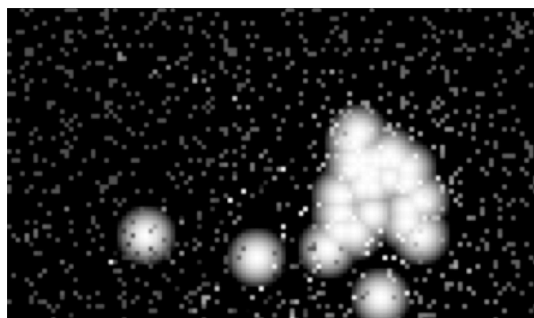


mancanti senza distinguere le mode identificate in corrispondenza della zona di fusione da quelle identificate in corrispondenza delle scintille, si assegnano luminosità elevate anche in corrispondenza dei pixel di *background* che sono adiacenti alle mode relative alle scintille. In questo modo, la ricostruzione che si ottiene in corrispondenza delle scintille, caratterizzate da una forma oblunga, non è rappresentativa della realtà. Inoltre, la mancata distinzione tra le mode relative alle zone di fusione regolare e quelle relative alle scintille avrà impatto anche sulla successiva fase di campionamento.

Seppure la luminosità delle mode identificate in corrispondenza delle scintille sia elevata tanto quanto quella delle altre mode, i pixel adiacenti alle scintille sono pixel di *background* e avranno luminosità non elevata. Allo scopo di distinguere le mode identificate in corrispondenza delle scintille dalle altre, si valutano dunque le luminosità acquisite in corrispondenza dei pixel adiacenti alle mode identificate all'istante temporale corrente. Dagli studi di simulazione condotti, si suggerisce di valutare le luminosità dei pixel che si trovano ad una distanza  $d$  dalla moda inferiore o uguale a 2. Qualora la luminosità osservata sia inferiore al 30% della luminosità della moda, si dichiara che quest'ultima è relativa ad una scintilla, eliminandola dall'insieme delle mode identificate all'istante temporale corrente. La valutazione della luminosità dei pixel adiacenti alle mode viene condotta per



FIGURA 4.6: Video 2 - frame 1

FIGURA 4.7: Ricostruzione del frame ( $h_1 = 7$ ).FIGURA 4.8: Ricostruzione del frame rimuovendo le mode relative alle scintille ( $h_1 = 7$ ).

tutte le mode identificate all'istante corrente, eliminando quelle relative alle scintille, ottenendo così un nuovo insieme di mode con numerosità ridotta. Successivamente, le misurazioni mancanti vengono imputate sulla base del nuovo insieme di mode. La ricostruzione del *frame* originale rimuovendo, dall'insieme delle mode, quelle relative alle scintille è riportata nella Figura 4.8.

Nonostante la ricostruzione nella Figura 4.8 sia più realistica di quella riportata nella Figura 4.7, come anticipato in precedenza, le scintille e la loro diffusione costituiscono una "firma" del processo e forniscono informazioni sullo stato stesso del processo (Colosimo et al., 2022). Oltre alla diffusione spaziale delle scintille, anche la loro forma e la loro dimensione possono fornire informazioni rilevanti riguardo la stabilità del processo e gli eventuali difetti che possono verificarsi durante la fase di produzione. Identificando le mode in corrispondenza delle scintille e distinguendole dalle altre, invece di rimuoverle dall'insieme delle mode senza imputare le misurazioni mancanti adiacenti ad esse, le scintille stesse potrebbero essere ricostruite insieme alla loro forma e alla loro dimensione, fornendo così ulteriori informazioni sulla stabilità del processo.

Pertanto, dato il contenuto informativo delle scintille emesse durante la fase di produzione, nei paragrafi seguenti si sceglie di imputare le misurazioni mancanti secondo la procedura descritta al paragrafo 3.2.2, senza rimuovere dall'insieme delle mode identificate in ogni istante temporale quelle relative alle scintille.

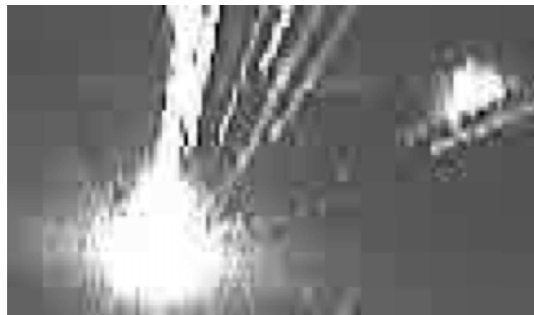
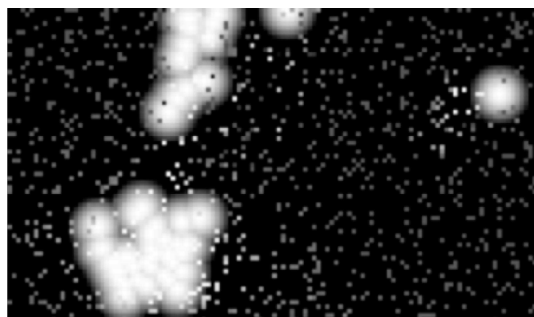
### 4.1.3 Regioni luminose multiple

La procedura di imputazione delle misurazioni mancanti e la strategia di campionamento proposte vengono applicate al *frame* nella Figura 4.9, che contiene prevalentemente due regioni luminose (la zona di fusione regolare e la zona affetta dall'*hotspot*), il raggio laser e le scintille. Osservando  $m = 1000$  pixel, selezionati in maniera casuale dal *frame*, si imputano le misurazioni mancanti ottenendo la ricostruzione riportata nella Figura 4.10. Applicando la procedura di imputazione delle misurazioni mancanti, entrambe le regioni luminose vengono ricostruite, nonostante la regione affetta dall'*hotspot* sia sparsa ed i pixel osservati siano stati selezionati in maniera casuale.

In corrispondenza di ciascun pixel che compone il *frame*, si assegna una probabilità di entrare a far parte del campione di  $m$  pixel che sarà osservato all'istante temporale

successivo, impiegando il *kernel*  $t$  di Student. Dagli studi di simulazione condotti, si suggerisce di selezionare un valore dell'ampiezza di banda  $h_2$  pari a 60. Tale valore è approssimativamente pari alla metà del numero di colonne di ciascun *frame*, estratto dai video disponibili, e sarà impiegato anche nei paragrafi successivi, relativi all'identificazione degli *hotspot*. Gli  $m$  pixel selezionati dalla strategia di campionamento, che verranno osservati all'istante successivo, sono rappresentati nella Figura 4.11. Con la selezione effettuata, i pixel appartenenti alle regioni luminose saranno osservati all'istante temporale successivo. In particolare, saranno osservate anche le luminosità dei pixel corrispondenti al raggio laser, poiché alcune mode sono state identificate in corrispondenza di quest'ultimo.

La risoluzione delle immagini non è elevata e, nel caso presente, avendo diverse zone luminose - anche particolarmente estese - all'interno del *frame* considerato, vengono identificate molte mode. In questo caso, la strategia di campionamento alloca la maggior parte degli  $m$  pixel in favore dello sfruttamento delle risorse, selezionando la maggior parte dei pixel in corrispondenza delle due regioni luminose e del raggio laser. Sfortunatamente, in questo caso, la strategia di campionamento tende ad assegnare poche risorse per l'esplorazione.

FIGURA 4.9: Video 1 - *frame* 8FIGURA 4.10: Ricostruzione del *frame* ( $h_1 = 7$ ).FIGURA 4.11: Pixel campionati al tempo successivo ( $h_2 = 60$ ).

#### 4.1.4 Regione sparsa

Per valutare sia la ricostruzione ottenuta in corrispondenza di un *frame* contenente un'unica regione luminosa sparsa, sia la selezione effettuata dalla strategia di campionamento, si applica l'approccio proposto al *frame* della Figura 4.12.

Selezionando da questo *frame*  $m = 1000$  pixel, in maniera casuale, si identificano due mode e si imputano le misurazioni mancanti ottenendo la ricostruzione riportata nella Figura 4.13. Nonostante le dimensioni della regione luminosa siano ridotte, la ricostruzione ottenuta con  $h_1 = 7$ , sulla base dei pixel campionati casualmente, è molto vicina a quella del *frame* originale.

Gli  $m$  pixel, osservati nell'istante temporale successivo, sono rappresentati nella Figura 4.14 e si concentrano principalmente nella zona del *frame* che comprende la regione sparsa. La regione luminosa, in questo caso, corrisponde ad un angolo acuto della porzione triangolare in cui si verifica un *hotspot*. I *frame* successivi catturano il lento raffreddamento della zona affetta dall'*hotspot* e, continuando ad osservare i pixel in questa zona del *frame*, è possibile valutare la loro luminosità e la variazione della luminosità nel tempo, nonostante le dimensioni della regione luminosa si riducano.



FIGURA 4.12: Video 1 - *frame* 9



FIGURA 4.13: Ricostruzione del *frame* ( $h_1 = 7$ ).

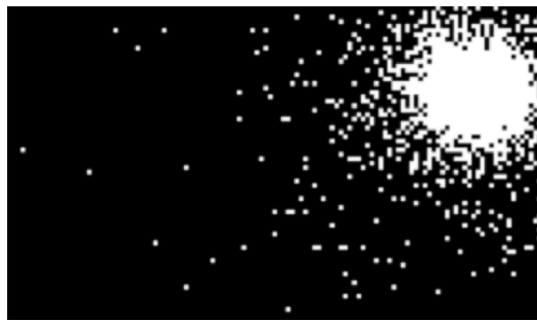


FIGURA 4.14: Pixel campionati al tempo successivo ( $h_2 = 60$ ).

### 4.1.5 Sequenza di *frame*

Estraendo un *frame* al secondo dal primo video dei tre disponibili, i *frame* riportati nelle Figure 4.9 e 4.12 vengono acquisiti sequenzialmente. Il tasso di acquisizione dei *frame* non è particolarmente elevato ma, estraendo un *frame* al secondo, è possibile apprezzare repentine variazioni della luminosità dei pixel.

L'approccio proposto viene applicato ai due *frame* estratti sequenzialmente. Inizialmente si osservano  $m$  pixel dal primo *frame*, estratti in maniera casuale. La ricostruzione del primo *frame* e la selezione effettuata dalla strategia di campionamento sono riportate nel paragrafo 4.1.3, rispettivamente, nelle Figure 4.10 e 4.11. Dal secondo *frame*, verranno osservati i pixel selezionati dalla strategia di campio-

namento e, sulla base delle misurazioni osservate, in corrispondenza di pixel adiacenti si identificano due mode e si imputano le misurazioni mancanti ottenendo la ricostruzione nella Figura 4.15. Seppure la maggior parte dei pixel sia stata allocata dalla strategia di campionamento in corrispondenza delle zone luminose, con le mode identificate a partire dalle misurazioni acquisite dal secondo *frame*, la ricostruzione ottenuta è soddisfacentemente vicina al *frame* originale.

Successivamente, in corrispondenza di ciascun pixel si assegna una probabilità di entrare a far parte del campione osservato all'istante temporale successivo e gli  $m$  pixel che saranno osservati all'istante successivo sono rappresentati nella Figura 4.16. La selezione effettuata si concentra nella zona del *frame* contenente la regione sparsa, esplorando anche nelle zone confinanti.

Nel paragrafo successivo, vengono riportati i risultati ottenuti applicando il metodo



FIGURA 4.15: Ricostruzione del secondo *frame* della sequenza ( $h_1 = 7$ ).



FIGURA 4.16: Pixel campionati al tempo successivo ( $h_2 = 60$ ).

di imputazione e la strategia di campionamento proposti al variare del numero  $m$  di pixel osservabili da ciascun *frame*.

## 4.2 Identificazione dei brillamenti solari

L'approccio proposto è applicato a *frame* consecutivi estratti dalla sequenza raccolta dal *Solar Dynamics Observatory*. La risoluzione dei *frame* è elevata, ciascuno ha dimensione  $292 \times 232$  ed è quindi composto da  $p = 67744$  pixel. Il metodo di imputazione e la strategia di campionamento vengono applicati supponendo, dapprima, di poter osservare solo il 10% dei pixel totali a causa della limitata banda di trasmissione dei dati e delle limitate risorse computazionali per l'analisi in tempo reale. In seguito, si applica l'approccio proposto assumendo di poter osservare, come in Liu et al. (2015), solo il 2.95% dei pixel che compongono ciascun *frame*, osservando dunque solo  $m = 2000$  pixel dei  $p$  totali.

### 4.2.1 Imputazione e campionamento: 10% dei pixel totali

L'approccio proposto viene applicato ai *frame* 217 e 218 della sequenza originale raccolta dal satellite, riportati rispettivamente nelle Figure 4.17 e 4.21. Dal *frame* 217 si osserva un brillamento solare che coinvolge un numero di pixel estremamente ridotto, la cui luminosità è intensa e rimane elevata anche nel *frame* 218 ed in quelli successivi.

Inizialmente, si campionano  $m = 6774$  pixel tra  $p$  totali, osservando il 10% dei pixel che compongono il *frame* (Figura 4.18). Sulla base delle misurazioni osservate, si identificano due mode ed impiegando  $h_1 = 5$  si imputano le misurazioni mancanti, ottenendo la ricostruzione riportata nella Figura 4.19. Il valore dell'ampiezza di banda è stato selezionato misurando le dimensioni del più piccolo rettangolo che circonda la regione luminosa del *frame* che precede la sequenza [217, 218], ossia il *frame* 216, riportato nella Figura 3.3. Si noti che, nella scelta dell'ampiezza di banda  $h_1$ , non è necessario selezionare un *frame* che contiene un brillamento solare e misurare le dimensioni di quest'ultimo, ma si suggerisce di selezionare un *frame* che contiene una regione luminosa di dimensioni medie, che non necessariamente rappresenta un andamento indesiderato o fuori controllo.

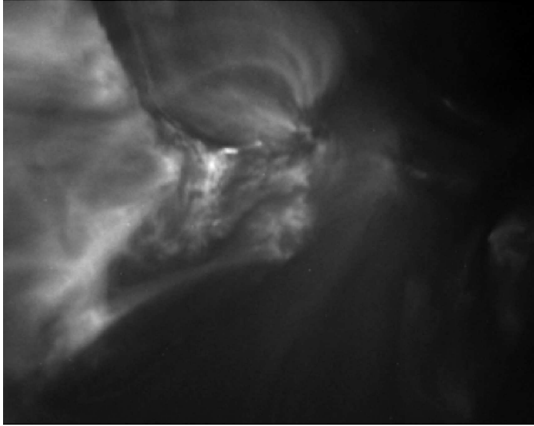


FIGURA 4.17: *Frame 217.*

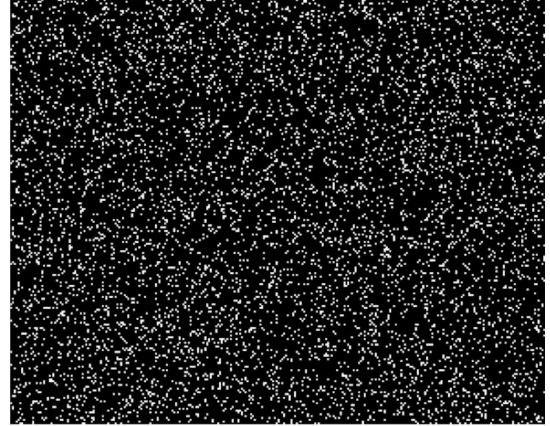


FIGURA 4.18: Campionamento casuale di  $m = 6774$  pixel dal *frame 217.*

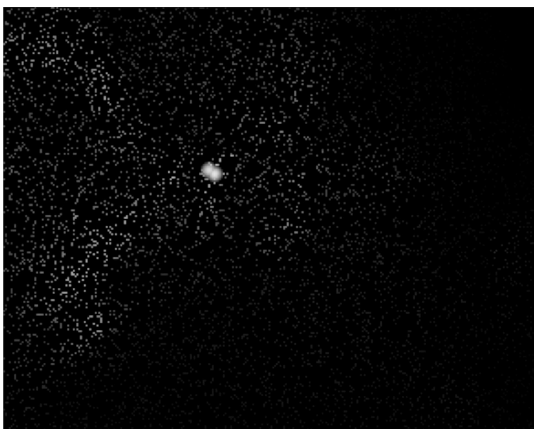


FIGURA 4.19: Ricostruzione del *frame 217* ( $h_1 = 5$ ).

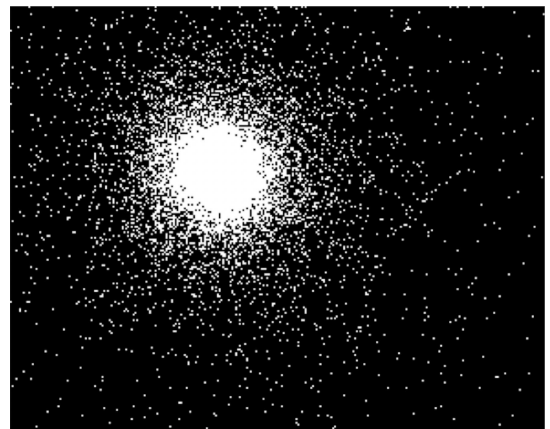


FIGURA 4.20: Campionamento di  $m = 6774$  pixel dal *frame 218* ( $h_2 = 150$ ).

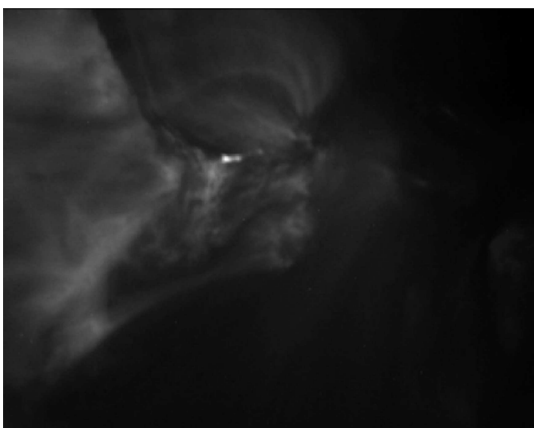


FIGURA 4.21: *Frame 218.*

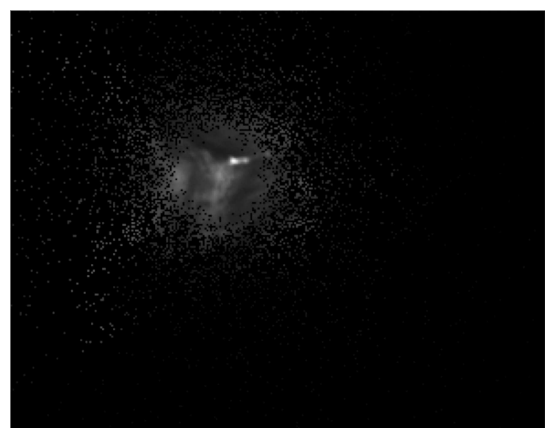


FIGURA 4.22: Ricostruzione del *frame 218* ( $h_1 = 5$ ).

Gli  $m$  pixel selezionati dalla strategia di campionamento, che saranno osservati in corrispondenza del *frame* 218, sono rappresentati nella Figura 4.20. Il valore dell'ampiezza di banda  $h_2$  è stato selezionato sulla base delle dimensioni dei *frame*, scegliendo  $h_2 = 150$ . Tale valore è approssimativamente pari alla metà del numero di colonne dei *frame* acquisiti e, seppure si tratti di un valore approssimato, la selezione effettuata dalla strategia di campionamento non differisce particolarmente da quella riportata nella Figura 4.20, quando si impiega un valore di  $h_2$  elevato. In questo caso, si consiglia di selezionare un valore di  $h_2$  maggiore di 100.

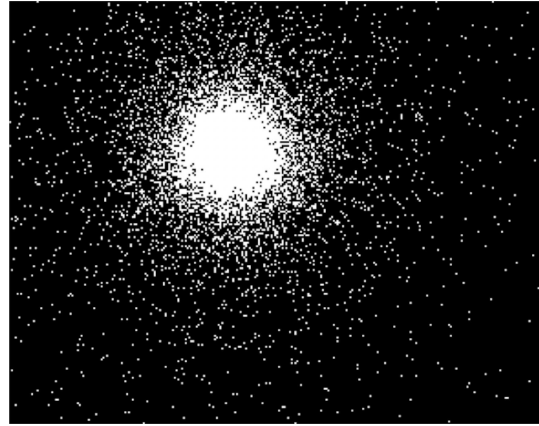


FIGURA 4.23: Campionamento di  $m = 6774$  pixel dal *frame* successivo della sequenza ( $h_2 = 150$ ).

In corrispondenza del *frame* 218, osservando gli  $m$  pixel selezionati dalla strategia di campionamento, si identificano le mode e si imputano le misurazioni mancanti ottenendo la ricostruzione riportata nella Figura 4.22. Pur avendo identificato due mode, sono state acquisite le luminosità dei pixel aventi una distanza  $d$ , dalla moda più vicina, inferiore ad  $h_1 = 5$ . Non vi sono, dunque, misurazioni mancanti attorno alle mode identificate e la regione luminosa del *frame* 218, estremamente sparsa, viene osservata interamente. Il valore selezionato per l'ampiezza di banda  $h_1$  ha un forte impatto sulle ricostruzioni ottenute quando i pixel vengono selezionati in maniera casuale; diversamente, quando la strategia di campionamento seleziona i pixel campionando in una o più regioni sospette e le regioni luminose sono estremamente sparse, il valore di  $h_1$  non ha un forte impatto sulle ricostruzioni ottenute.

I pixel campionati dalla strategia, che verranno osservati nel *frame* successivo della sequenza, sono rappresentati nella Figura 4.23. La selezione dei pixel osservati nel *frame* successivo si basa sulle mode identificate nel *frame* corrente e la zona nella quale si concentra la maggior parte dei pixel campionati è centrata attorno alle mode identificate. *Frame* dopo *frame*, tale zona si sposta, centrandosi in corrispondenza della regione luminosa osservata in corrispondenza del *frame* corrente.



### 4.2.2 Imputazione e campionamento: 2.95% dei pixel totali

Impiegando i *frame* 217 e 218, l'approccio proposto è applicato supponendo di poter osservare, da ogni *frame* della sequenza, solo  $m = 2000$  pixel che corrispondono al 2.95% dei pixel che compongono ciascun *frame*.

Dal *frame* 217, si osservano le luminosità di  $m$  pixel, campionati in maniera casuale (Figura 4.24). Si identificano due mode e si imputano le misurazioni mancanti impiegando  $h_1 = 5$ ; la ricostruzione del *frame* 217, ottenuta con il metodo di imputazione proposto, è riportata nella Figura 4.25.

In corrispondenza di ciascun pixel, sulla base della sua distanza dalla moda più vicina, si assegna una probabilità di entrare a far parte del campione di  $m$  pixel che sarà osservato al *frame* successivo. Come in precedenza, si impiega un valore dell'ampiezza di banda pari a  $h_2 = 150$ . I pixel campionati dal *frame* 218 sono rappresentati nella Figura 4.26. Osservando gli  $m$  pixel, si identifica un'unica moda e si imputano le misurazioni mancanti ottenendo la ricostruzione nella Figura 4.27. In questo caso, sono state acquisite le luminosità dei pixel aventi una distanza  $d$  dalla moda inferiore ad  $h_1 = 5$ . La regione luminosa è

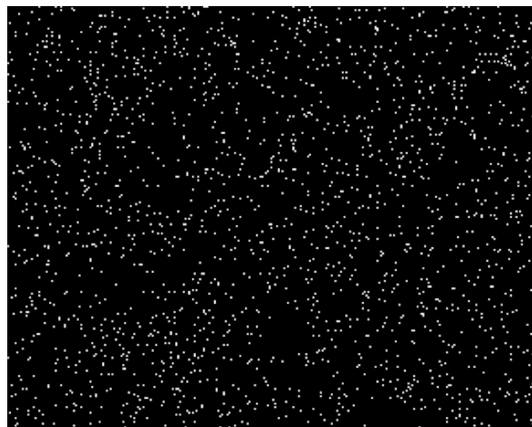


FIGURA 4.24: Campionamento casuale di  $m = 2000$  pixel dal *frame* 217.



FIGURA 4.25: Ricostruzione del *frame* 217 ( $h_1 = 5$ ).

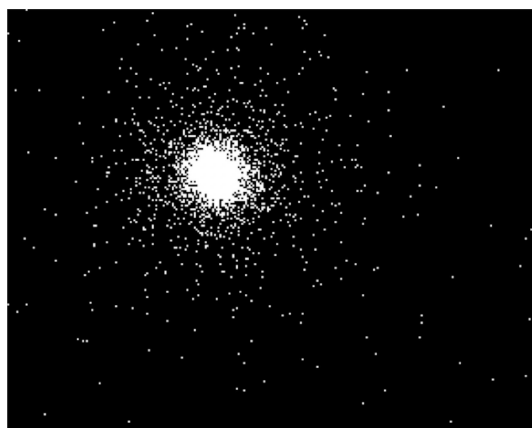


FIGURA 4.26: Campionamento di  $m = 2000$  pixel dal *frame* 218 ( $h_2 = 150$ ).

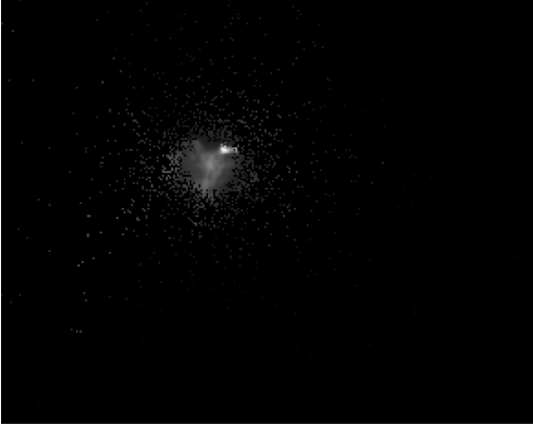


FIGURA 4.27: Ricostruzione del *frame* 218 ( $h_1 = 5$ ).

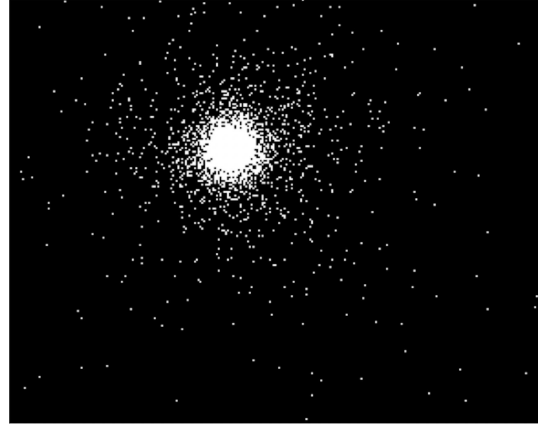


FIGURA 4.28: Campionamento di  $m = 2000$  pixel dal *frame* successivo della sequenza ( $h_2 = 150$ ).

estremamente sparsa e, con la selezione effettuata, tale regione, che corrisponde al brillamento solare, viene osservata quasi completamente nonostante si osservi solo il 2.95% dei pixel che compongono il *frame*.

Gli  $m$  pixel selezionati dalla strategia di campionamento, le cui luminosità saranno osservate nel *frame* successivo della sequenza, sono rappresentati nella Figura 4.28. Come descritto in precedenza, la selezione dei pixel da osservare nel *frame* successivo si basa sulle mode identificate nel *frame* disponibile all'istante corrente. In questo caso, è stata identificata un'unica moda. La zona del *frame*, nella quale si concentra la maggior parte dei pixel campionati, si sposta centrandosi in corrispondenza della moda e, dunque, della regione luminosa osservata nel *frame* disponibile all'istante temporale corrente.

La strategia di campionamento sfrutta la vicinanza spaziale dei pixel per assegnare, in corrispondenza di ciascuno di essi, una probabilità di essere selezionato al tempo successivo. Assegnando le probabilità, sulla base della distanza di ciascun pixel dalla moda più vicina, la strategia di campionamento tende a favorire più lo sfruttamento che l'esplorazione delle risorse. Con l'obiettivo di assegnare un numero maggiore di risorse per l'esplorazione, altre tipologie di funzioni *kernel* potrebbero essere impiegate dalla strategia di campionamento. Similmente, si potrebbe valutare l'impatto della scelta di altre tipologie di funzioni *kernel* sulle ricostruzioni dei *frame*.

# Conclusioni

Tramite l'analisi in tempo reale di immagini e di video, acquisiti ad elevata frequenza, la sorveglianza di un processo può essere condotta analizzando la luminosità e la variazione della luminosità nel tempo. In numerose applicazioni, tuttavia, sono disponibili solo osservazioni parziali, in ogni istante temporale, per valutare la stabilità del processo. Diverse metodologie di sorveglianza sequenziale sono state sviluppate assumendo che solo un sottoinsieme di pixel sia osservabile da ciascuna immagine acquisita. In particolare, metodologie parametriche e non parametriche sono state sviluppate per ricostruire le misurazioni mancanti, sorvegliare la stabilità del processo e selezionare il sottoinsieme di pixel da osservare, all'istante temporale successivo, in assenza di allarme. Alcune di queste impiegano l'algoritmo *Thompson Sampling* come strategia di campionamento, che seleziona il sottoinsieme di pixel da osservare, bilanciando il compromesso tra l'esplorazione e lo sfruttamento delle risorse, in un contesto di risorse limitate.

Con l'obiettivo di proporre un nuovo approccio per ricostruire le misurazioni mancanti e campionare il sottoinsieme di pixel da osservare, nella presente trattazione, si illustra una procedura non parametrica che, a differenza della maggior parte delle metodologie sviluppate, non richiede siano soddisfatti particolari assunti distributivi difficili da soddisfare nelle applicazioni reali. L'approccio impiega le luminosità acquisite nel loro formato originale ed identifica, in ogni istante temporale, le mode della distribuzione sottostante. Identificando queste ultime e sfruttando la vicinanza spaziale dei pixel, le misurazioni mancanti vengono imputate, fornendo una ricostruzione del *frame* originale. Le mode identificate e la vicinanza spaziale, di ciascun pixel dalla moda più vicina, vengono impiegate dalla strategia di campionamento proposta per assegnare, ad ogni pixel, una probabilità di essere osservato nel *frame* successivo. La strategia di campionamento è in grado di selezionare pixel appartenenti a regioni multiple dello spazio, nelle quali

l'andamento diviene sospetto ed indesiderato. La strategia implementata tende ad assegnare la maggior parte dei pixel nelle regioni sospette, favorendo più lo sfruttamento che l'esplorazione delle risorse. Altre tipologie di funzioni *kernel* potrebbero però essere impiegate dalla strategia di campionamento, allo scopo di assegnare un numero maggiore di pixel per l'esplorazione. Inoltre, si potrebbe valutare anche l'impatto della scelta di altre tipologie di funzioni *kernel* sulle ricostruzioni dei *frame*, ottenute applicando l'approccio proposto.

La procedura presentata è applicata a due casi reali, riferiti a due contesti applicativi molto distanti tra loro. In particolare, il primo riguarda le tecnologie di fabbricazione additiva e l'identificazione dei difetti durante la fase di produzione, mentre il secondo riguarda l'osservazione di fenomeni fisici che possono danneggiare le infrastrutture tecnologiche. Altri ambiti applicativi, differenti da quelli citati, sono caratterizzati dalla disponibilità di immagini parziali, alle quali potrebbe essere applicato l'approccio proposto. Ad esempio, con l'obiettivo di prevenire il crimine, l'approccio potrebbe essere applicato per la ricostruzione di *frame*, parzialmente osservabili, acquisiti dalle telecamere di sicurezza. Inoltre, il metodo di imputazione delle misurazioni mancanti e la strategia di campionamento proposti potrebbero essere abbinati ad uno schema di controllo per sorvegliare un processo in presenza di osservazioni parziali. Senza limitarsi alla sorveglianza di flussi di immagini, l'approccio potrebbe poi essere applicato in numerosi contesti in cui sono tipicamente disponibili solo osservazioni parziali, provenienti da sensori disposti in una specifica area geografica. Ad esempio, per sorvegliare i livelli di inquinamento dell'aria, delle acque o la temperatura del suolo, quando i sensori disponibili non possono essere mantenuti costantemente in funzione.

# Appendice

L'applicazione del metodo di imputazione delle misurazioni mancanti e della strategia di campionamento ai due casi reali può essere riprodotta implementando il codice R riportato in seguito.

I video registrati allo scopo di sorvegliare il processo SLM sono disponibili all'indirizzo <http://doi.org/10.6084/m9.figshare.7092863>. I *frame* della sequenza raccolta dal *Solar Dynamics Observatory* sono disponibili all'indirizzo <http://nislab.ee.duke.edu/MOUSSE/index.html>; oppure, in alternativa, è disponibile un file MAT, contenente la sequenza di *frame* raccolta dal satellite, nel materiale supplementare della metodologia TS-BSSCD (Guo et al., 2022). Per estrarre i *frame* dai video e dal file MAT sono state impiegate le funzioni riportate in seguito le quali, a loro volta, utilizzano delle funzioni contenute nei pacchetti R: *av* (Ooms, 2023), *EImage* (Pau et al., 2010) e *R.matlab* (Bengtsson, 2022).

```
### Funzione per estrarre i frame da un video e salvarli in una
      cartella
from_video <- function(path_video, path_dest, format = "jpg",
      fps=NULL){
  # path_video: indirizzo del video
  # path_dest: indirizzo della cartella
  setwd(path_dest)
  out <- paste0("image_%d.", format) # output
  a_rate <- paste0("fps=fps=", fps) # acquisition rate
  av::av_encode_video(input = path_video, output = out, vfilter =
    a_rate)
}
```

```
### Funzione per creare un array contenente frame in scala di grigi
frames_to_array <- function(path_dest, nframes){
  # path_dest: indirizzo della cartella che contiene i frame
  # nframes: numero di frame da estrarre
  require(EBImage)
  frame1 <- readImage(paste0(path_dest, "/image_1.jpg"))
  frame1 <- channel(frame1, "gray")
  # preallocazione dell'array
  frames <- array(NA, dim = c(dim(frame1)[1], dim(frame1)[2],
    nframes))
  for(i in 1:nframes){
    f <- readImage(paste0(path_dest, "/image_", i, ".jpg"))
    frames[, ,i] <- channel(f, "gray")
  }
  return(frames = frames)
}
```

```
### Funzione per creare un array contenente frame in scala di grigi
  estratti da un file MAT
d <- R.matlab::readMat(path_data)
# path_data: indirizzo del file .mat
from_Mat <- function(dati){
  my_frames <- array(NA, dim = c(dim(dati$data)[1],
    dim(dati$data)[2], dim(dati$data)[3]))
  nframes <- dim(dati$data)[3]
  for(i in 1:nframes){
    one_frame <- dati$data[, ,i]
    one <- as.vector(one_frame)
    # normalizzazione per avere gray-scale:
    normfr <- (one-min(one))/(max(one)-min(one))
    my_frames[, ,i] <- matrix(data = normfr, nrow=dim(my_frames)[1],
      ncol = dim(my_frames)[2])
  }
  return(frames = my_frames)
}
```

Dopo aver estratto i *frame*, salvandoli in un array chiamato `frames`, si costruisce una matrice, chiamata `pixel_coords`, contenente le coordinate di colonna e di riga di ogni pixel che compone ciascun *frame*. Si imposta, inoltre, il numero di pixel  $m$  che si assume siano osservabili da ciascun *frame*, acquisito in ogni istante temporale.

Con la funzione `distance_mat` si calcola la matrice delle distanze euclidee, di dimensione  $p \times p$ , tra tutti i pixel che compongono ciascun *frame*. La matrice può essere calcolata *offline* ed un'unica volta poiché le distanze tra ciascun pixel e tutti gli altri sono fissate ed indipendenti dalla luminosità e dalla variazione della luminosità registrata in corrispondenza di ciascun pixel. Le distanze euclidee verranno impiegate, nel seguito, nel calcolo della distanza tra ciascun pixel e la moda più vicina ad esso. Talvolta, il numero di pixel che compone ciascun *frame* è molto elevato e non è possibile allocare in memoria una matrice di dimensione  $p \times p$ . Le distanze euclidee tra i pixel, tuttavia, possono essere calcolate anche *online*. Di seguito, si riporta una funzione alternativa per il calcolo della distanza tra ciascun pixel e la moda più vicina, da implementare nei casi in cui le distanze debbano essere calcolate *online*.

Al primo istante temporale, dal primo *frame* contenuto nell'array, si campionano casualmente  $m$  pixel e si identificano le mode, sulla base delle  $m$  misurazioni osservate, con la funzione `find_modes` (cfr. paragrafo 3.2.1). La funzione prende in input (`v`) una matrice  $p \times 4$  che contiene in ogni riga, ossia per ogni pixel: la sua etichetta (che assume valori da  $1, \dots, p$ ), la coordinata di colonna, la coordinata di riga e la luminosità osservata oppure mancante. Le righe della matrice sono ordinate, in maniera decrescente, rispetto alle luminosità. La funzione `find_modes` restituisce in output una matrice avente numero di righe pari al numero di mode identificate e numero di colonne pari a 4, riportando, per ogni moda identificata, la sua etichetta, le coordinate di colonna e di riga e la luminosità della moda.

```
dim1 <- dim(frames)[1] # numero di colonne
dim2 <- dim(frames)[2] # numero di righe
pixel_coords <- expand.grid(x1 = 1:dim1, x2 = 1:dim2)
p <- dim(pixel_coords)[1]
# p: numero di pixel che compone ciascun frame
# pixel_coords e' una matrice p x 2 che contiene le coordinate di
  colonna e di riga di ciascun pixel
```

```

m <- floor(0.1*p) # pixel osservabili

### distance_mat (fun)
distance_mat <- function(pixel_coords){
  x1 <- pixel_coords[,1]
  x2 <- pixel_coords[,2]
  diff_x1 <- outer(x1, x1, "-")
  diff_x2 <- outer(x2, x2, "-")
  dist_mat <- sqrt(diff_x1^2 + diff_x2^2)
  return(dist_mat)
}

### Random Sampling
W <- rep(1/p, p)
set.seed(42)
sampled.pix <- sample(1:p, size = m, prob = W)
coord.sampled <- pixel_coords[sampled.pix,]
new_image <- rep(NA,p)
# osserviamo gli m pixel selezionati casualmente dal frame 1:
new_image[sampled.pix] <- as.vector(frames[, ,1])[sampled.pix]
really_obs = cbind(pixel_coords, new_image)
# really_obs e' una matrice p x 3; la terza colonna contiene le m
  luminosita' osservate e le p-m mancanti (NA)

### sampled pixels: forprint (fun)
forprint <- function(really_obs = really_obs){
  to_plot <- rep(NA, nrow(really_obs))
  for(i in 1:nrow(really_obs)){
    to_plot[i] <- ifelse(is.na(really_obs[i,3]),0, 1)
  }
  return(to_plot = cbind(really_obs, to_plot))
}

stamp <- forprint(really_obs)
to_print <- matrix(stamp[,4], nrow=dim1, ncol = dim2)
# graficamente guardiamo quali pixel abbiamo campionato con il
  sampling:
display(to_print, method = "raster")

```



```

### Identificazione delle mode: find_modes (fun)
etichette <- order(really_obs[,3], decreasing = T)
# etichette ordinate in ordine decrescente rispetto alle luminosita'
ord_vec <- really_obs[etichette,]
possible.m <- cbind(etichette,ord_vec)
# possible.m: input delle funzioni find_modes, dist_fmB, dist_withoutm

find_modes <- function(v){
  # v (p x 4) contiene in ogni riga: etichetta, coordinata di
  # colonna, coordinata di riga e luminosita' osservata oppure NA
  modes <- data.frame(matrix(NA, nrow = nrow(v), ncol = 4))
  colnames(modes) <- colnames(v)
  if(v[2,4]-v[3,4] > v[1,4]-v[2,4]){
    modes[1,] <- v[1,]
    modes[2,] <- v[2,]
  } else if(v[2,4]-v[3,4] < v[1,4]-v[2,4]){
    modes[1,] <- v[1,]
  } else {
    count = 1
    while((v[count,4]-v[count+1,4]) >= (v[count+1,4]-v[count+2,4]) &
count <= (p-2)){
      modes[count,] <- v[count,]
      modes[count+1,] <- v[count+1,]
      modes[count+2,] <- v[count+2,]
      count = count + 1
    }
  }
  modes <- na.omit(modes)
  return(modes = modes)
}

```

Dopo aver identificato le mode, si calcola la distanza tra ciascun pixel e la moda più vicina con la funzione `dist_fmB`. Oltre alla distanza, viene memorizzata anche la luminosità della moda più vicina. Quando l'elevata dimensionalità non permette l'allocatione in memoria della matrice delle distanze euclidee è possibile impiegare, in alternativa, la funzione `dist_withoutm`. L'output restituito da una delle due funzioni alternative viene passato in input, assieme all'ampiezza di banda  $h_1$ , alla funzione `KImpute`, con

la quale si imputano le misurazioni mancanti. In particolare, oltre ad  $h_1$ , la funzione `KImpute` riceve in input (`object`) una matrice di dimensione  $p \times 6$  contenente in ogni riga, ossia per ciascun pixel: l'etichetta del pixel, le coordinate di colonna e di riga, la luminosità osservata oppure mancante, la distanza dalla moda più vicina e la luminosità della moda più vicina. In output, la funzione `KImpute` restituisce una matrice  $p \times 7$  ottenuta, concatenando in colonna, la matrice in input con il vettore contenente le  $m$  luminosità osservate e le  $p-m$  luminosità imputate con la procedura proposta, descritta nel paragrafo 3.2.2.

Con riferimento ai video registrati per la sorveglianza del processo SLM e ai *frame* estratti da questi, per riconoscere le mode relative alle scintille, rimuoverle dall'insieme delle mode identificate all'istante temporale corrente ed imputare le misurazioni mancanti, in alternativa alle funzioni `dist_fmB` e `KImpute`, si impiegano le funzioni `dist_fmB_sp` e `Impute_nosp` (cfr. paragrafo 4.1.2).

```
### Calcolo della distanza dalla moda piu' vicina: dist_fmB (fun)
dist_fmB <- function(modes, distmat, possible.m){
  # modes: output di find_modes
  # distmat: matrice delle distanze calcolata con distance_mat (fun)
  p <- nrow(possible.m)
  dis <- matrix(NA, nrow = p, ncol = nrow(modes))
  br <- matrix(NA, nrow = p, ncol = 1)
  for(i in 1:ncol(dis)){
    for(j in 1:nrow(dis)){
      m1 <- modes[i,1] # etichetta di ciascuna moda
      m2 <- possible.m[j,1] # etichetta di ciascun pixel
      if(m1 != m2){
        dis[j, i] <- distmat[m1 ,m2]
      } else if(m1 == m2){
        dis[j, i] <- 0
      }
    }
  }
  br <- possible.m$new_image[apply(dis,1,which.min)]
  # br: brightness della moda piu' vicina
  dis <- apply(dis,1,min)
```

```

# dis: distanza dalla moda piu' vicina
tab <- cbind(possible.m, dis, br)
return(dist.fm = tab)
}

### Calcolo della distanza dalla moda piu' vicina senza richiedere in
input distmat: dist_withoutm (fun)
dist_withoutm <- function(modes, possible.m){
  p <- nrow(possible.m)
  dis <- matrix(NA, nrow = p, ncol = nrow(modes))
  br <- matrix(NA, nrow = p, ncol = 1)
  for(i in 1:ncol(dis)){
    for(j in 1:nrow(dis)){
      m1 <- modes[i,1] # etichetta di ciascuna moda
      m2 <- possible.m[j,1] # etichetta di ciascun pixel

      if(m1 != m2){
        x1_m1 <- possible.m[etichette==m1,2]
        x2_m1 <- possible.m[etichette==m1,3]
        x1_m2 <- possible.m[etichette==m2,2]
        x2_m2 <- possible.m[etichette==m2,3]
        dis[j,i] <- (((x1_m1 - x1_m2)^2) + ((x2_m1 - x2_m2)^2))^0.5
      } else if(m1 == m2){
        dis[j, i] <- 0
      }
    }
  }
  br <- possible.m$new_image[apply(dis,1,which.min)]
  # br: brightness della moda piu' vicina
  dis <- apply(dis,1,min)
  # dis: distanza dalla moda piu' vicina
  tab <- cbind(possible.m, dis, br)
  return(dist.fm = tab)
}

### Kernel functions:
### Epanechnikov Kernel: epanK (fun)
epanK <- function(d, h){
  # d e' un vettore contenente la distanza di ciascun pixel dalla

```

```

# moda piu' vicina; h e' l'ampiezza di banda
ep_weight <- rep(NA, length(d))
K_Epa <- function(x, h){
  3/(4*h) * (1 - (x / h)^2) * (abs(x/h) <= 1)
}
h <- h
ep_weight <- sapply(d, K_Epa, h = h)
return(w = ep_weight)
}

### Student's t Kernel: studentK (fun)
studentK <- function(d, h){
  nu <- (2*(h^2))/((h^2)-1)
  if(nu <= 2)
    stop("Error")
  t_weight <- rep(NA, length(d))
  t_weight <- sapply(d, function(x) dt(x, df = nu))
  return(w = t_weight)
}

### Imputazione delle misurazioni mancanti: KImpute (fun)
KImpute <- function(object, h){
  # object: output di dist_fmB oppure di dist_withoutm
  # h: ampiezza di banda (h1)
  # object e' la matrice p x 6 che contiene per ogni pixel:
  # etichetta, coord. di colonna, coord. di riga, luminosita'
  # osservata/NA, distanza dalla moda piu' vicina, brightness della
  # moda piu' vicina
  p <- dim(object)[1]
  m <- length(na.omit(object$new_image))
  impute <- object$new_image
  impute[(m+1):p] <- (epanK(object$dis[(m+1):p],
    h=h)/epanK(0,h=h))*(object$br[(m+1):p])
  obj <- cbind(object, impute)
  return(fully = obj)
}

```

```
# per visualizzare la ricostruzione del frame: sia 'imputed' l'output
  restituito dalla funzione KImpute
printing <- cbind(imputed$etichette, imputed$impute)
# riordinare rispetto alle etichette:
printing <- printing[order(printing[,1]),]
to_print1 <- matrix(printing[,2], nrow=dim1, ncol = dim2)
# confronto grafico:
par(mfrow=c(1,2))
display(frames[,1], method = "raster")
display(to_print1, method = "raster")

### Riconoscimento, rimozione delle mode relative alle scintille ed
  imputazione delle misurazioni mancanti: dist_fmB_sp (fun) &
  Impute_nosp (fun)
dist_fmB_sp <- function(modes, distmat, possible.m, spatters = T){
  p <- nrow(possible.m)
  dis <- matrix(NA, nrow = p, ncol = nrow(modes))
  br <- matrix(NA, nrow = p, ncol = 1)
  for(i in 1:ncol(dis)){
    for(j in 1:nrow(dis)){
      m1 <- modes[i,1] # etichetta di ciascuna moda
      m2 <- possible.m[j,1] # etichetta di ciascun pixel
      if(m1 != m2){
        dis[j, i] <- distmat[m1 ,m2]
      } else if(m1 == m2){
        dis[j, i] <- 0
      }
    }
  }
}

if(spatters){
  nm <- possible.m$etichette[apply(dis,1,which.min)]
  # nm: etichetta della moda piu' vicina
  br <- possible.m$new_image[apply(dis,1,which.min)]
  # br: brightness della moda piu' vicina
  dis <- apply(dis,1,min)
  # dis: distanza dalla moda piu' vicina
  tab <- cbind(possible.m, dis, br, nm)
  return(dist.fm = tab)
```

```

} else {
  br <- possible.m$new_image[apply(dis,1,which.min)]
  dis <- apply(dis,1,min)
  tab <- cbind(possible.m, dis, br)
}
}

Impute_nosp <- function(object, h, modes, distmat, min.br, h_control){
  # min.br varia in (0, 1) e viene moltiplicata per la luminosita'
  # della moda piu' vicina a ciascun pixel, tra gli m osservati
  # (min.br*luminosita' della moda) e' la soglia con cui vengono
  # confrontate le luminosita' osservate
  # h_control: distanza dalla moda entro la quale si confrontano le
  # luminosita' osservate con la soglia
  p <- dim(object)[1]
  m <- length(na.omit(object$new_image))
  nmodes <- nrow(object[object$dis == 0,])
  spatt <- NULL
  for(i in (nmodes+1):m){
    one_pixel <- object[i,]
    if(one_pixel$dis <= h_control & one_pixel$new_image <=
min.br*one_pixel$br){
      # salviamo l'etichetta della moda relativa alla scintilla:
      spatt <- rbind(spatt, one_pixel$nm)
    }
  }
  spatt <- unique(spatt)
  if(length(spatt) > 0) {
    possible.m <- object[,1:4]
    real_modes <- modes[!(modes$etichette %in% spatt),]
    # real_modes: nuovo insieme (ridotto) di mode
    # ricalcola la distanza dalla moda piu' vicina:
    object <- dist_fmB_sp(modes = real_modes, distmat = distmat,
possible.m = possible.m, spatters = F)
  }
  impute <- object$new_image
  impute[(m+1):p] <- (epanK(object$dis[(m+1):p],
h=h)/epanK(0,h=h))*(object$br[(m+1):p])
}

```

```

  obj <- cbind(object, impute)
  return(fully = obj)
}

```

Per assegnare, in corrispondenza di ciascun pixel, una probabilità di entrare a far parte del campione di pixel che sarà osservato all'istante temporale successivo si impiega la funzione `Sampling`. La funzione riceve in input (`object`) la matrice di dimensione  $p \times 6$ , che viene passata in input anche alla funzione `KImpute`, riportata precedentemente. Inoltre, la funzione `Sampling` riceve in input (`h`) l'ampiezza di banda  $h_2$  e restituisce in output il vettore delle probabilità assegnate in corrispondenza di ciascun pixel all'istante temporale corrente (cfr. paragrafo 3.2.3). Tale vettore viene riordinato rispetto all'etichetta di ciascun pixel e viene impiegato per campionare gli  $m$  pixel da osservare all'istante temporale successivo o, equivalentemente, dal *frame* successivo appartenente alla sequenza o estratto dal video.

```

### Campionamento: Sampling (fun)
Sampling <- function(object, h){
  # object: come input di KImpute (fun)
  # h: ampiezza di banda (h2)
  p <- dim(object)[1]
  N <- sum(studentK(object$dis, h = h))
  # N: costante di normalizzazione
  W <- studentK(object$dis, h = h)/N
  return(W = W)
}

# sia 'W' l'output restituito dalla funzione Sampling ed 'imputed'
# l'output restituito dalla funzione KImpute
ord_p <- cbind(imputed$etichette, W)
# riordinare le etichette:
ord_p <- ord_p[order(ord_p[,1]),]
sampled.pix <- sample(1:p, size = m, prob = ord_p[,2])

```





# Bibliografia

- AGRAWAL, S. & GOYAL, N. (2012). Analysis of Thompson Sampling for the multi-armed bandit problem. In *Proceedings of the 25th Annual Conference on Learning Theory*, S. Mannor, N. Srebro & R. C. Williamson, eds., vol. 23 of *Proceedings of Machine Learning Research*. Edinburgh, Scotland: PMLR.
- AITKIN, M. (1991). Posterior bayes factors. *Journal of the Royal Statistical Society. Series B (Methodological)* **53**, 111–142.
- ALTMAN, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* **46**, 175–185.
- AUGUSTO, C. R. A., FAUTH, A. C., NAVIA, C. E., SHIGEOUKA, H. & TSUI, K. H. (2011). Connection among spacecrafts and ground level observations of small solar transient events. An overview of small solar transient events observed at ground. *Experimental Astronomy* **31**, 177–197.
- BELL, R. C., JONES-FARMER, L. A. & BILLOR, N. (2014). A distribution-free multivariate Phase I location control chart for subgrouped data from elliptical distributions. *Technometrics* **56**, 528–538.
- BENGTSSON, H. (2022). *R.matlab: Read and Write MAT Files and Call MATLAB from Within R*. R package version 3.7.0.
- BUGATTI, M. & COLOSIMO, B. M. (2021). Towards real-time in-situ monitoring of hot-spot defects in L-PBF: a new classification-based method for fast video-imaging data analysis. *Journal of Intelligent Manufacturing* **33**, 293–309.
- BUI, A. T. & APLEY, D. W. (2018a). A monitoring and diagnostic approach for stochastic textured surfaces. *Technometrics* **60**, 1–13.

- BUI, A. T. & APLEY, D. W. (2018b). Monitoring for changes in the nature of stochastic textured surfaces. *Journal of Quality Technology* **50**, 363–378.
- CAPIZZI, G. (2015). Recent advances in process monitoring: Nonparametric and variable-selection methods for Phase I and Phase II. *Quality Engineering* **27**, 44–67.
- CAPIZZI, G. & MASAROTTO, G. (2017). Phase I distribution-free analysis of multivariate data. *Technometrics* **59**, 484–495.
- CHAPELLE, O. & LI, L. (2011). An empirical evaluation of Thompson Sampling. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS'11. Red Hook, NY, USA: Curran Associates Inc.
- CHENG, C. & SHIAU, J.-J. (2014). A distribution-free multivariate control chart for Phase I applications. *Quality and Reliability Engineering International* **31**, 97–111.
- COLOSIMO, B. & GRASSO, M. (2018). Spatially weighted PCA for monitoring video image data with application to additive manufacturing. *Journal of Quality Technology* **50**, 391–417.
- COLOSIMO, B. M., PAGANI, L. & GRASSO, M. (2022). Modeling spatial point processes in video-imaging via Ripley's K-function: an application to spatter analysis in additive manufacturing. *Journal of Intelligent Manufacturing* , 1–19.
- EPANECHNIKOV, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications* **14**, 153–158.
- GITTINS, J. C. (1989). *Multi-armed bandit allocation indices*. Wiley interscience series in systems and optimization. Chichester: John Wiley & Sons.
- GRASSO, M., LAGUZZA, V., SEMERARO, Q. & COLOSIMO, B. M. (2017). In-Process Monitoring of Selective Laser Melting: Spatial Detection of Defects Via Image Data Analysis. *Journal of Manufacturing Science and Engineering* **139**, 051001.
- GRASSO, M., REMANI, A., DICKINS, A., COLOSIMO, B. & LEACH, R. (2021). In-situ measurement and monitoring methods for metal powder bed fusion: An updated review. *Measurement Science and Technology* **32**, 112001.

- GUO, J., YAN, H. & ZHANG, C. (2022). A bayesian partially observable online change detection approach with Thompson Sampling. *Technometrics* **65**, 179–191.
- GÓMEZ, A. M. E., LI, D. & PAYNABAR, K. (2022). An adaptive sampling strategy for online monitoring and diagnosis of high-dimensional streaming data. *Technometrics* **64**, 253–269.
- HASTIE, T. J., TIBSHIRANI, R., FRIEDMAN, J. & HASTIE, T. J. (2009). *The elements of statistical learning : data mining, inference, and prediction*. Springer series in statistics. New York: Springer, 2nd ed.
- HENNIG, C., MEILĂ, M., MURTAGH, F. & ROCCI, R. (2015). *Handbook of Cluster Analysis*. New York: Chapman and Hall/CRC, 1st ed.
- HOI, S. C., SAHOO, D., LU, J. & ZHAO, P. (2021). Online learning: A comprehensive survey. *Neurocomputing* **459**, 249–289.
- HONG, Y., LIAO, H., YASHCHIN, E. & TSUNG, F. (2018). Editor’s notes on special issue on “reliability and maintenance modeling with big data”. *Journal of Quality Technology* **50**, 133–134.
- JIANG, B., WANG, C.-C. & LIU, H.-C. (2005). Liquid crystal display surface uniformity defect inspection using analysis of variance and exponentially weighted moving average techniques. *International Journal of Production Research - INT J PROD RES* **43**, 67–80.
- JIANG, B. C. & JIANG, S. (1998). Machine vision based inspection of oil seals. *Journal of Manufacturing Systems* **17**, 159–166.
- JOLLIFFE, I. (2002). *Principal component analysis*. Springer series in statistics. New York: Springer, 2nd ed.
- KOLDA, T. G. & BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Review* **51**, 455–500.
- LAI, T. & ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics* **6**, 4–22.

- LATTIMORE, T. & SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge: Cambridge University Press.
- LIU, K., MEI, Y. & SHI, J. (2015). An adaptive sampling strategy for online high-dimensional process monitoring. *Technometrics* **57**, 305–319.
- LU, C.-J. & TSAI, D.-M. (2005). Automatic defect inspection for LCDs using singular value decomposition. *International Journal of Advanced Manufacturing Technology* **25**, 53–61.
- MANI, M., LANE, B., DONMEZ, M., FENG, S., MOYLAN, S. & FESPERMAN, R. (2015). Measurement science needs for real-time control of additive manufacturing powder bed fusion processes. In *NIST Interagency/Internal Report (NISTIR)*. Gaithersburg, Maryland: National Institute of Standards and Technology (NIST).
- MEGAHED, F. M., WOODALL, W. H. & CAMELIO, J. A. (2011). A review and perspective on control charting with image data. *Journal of Quality Technology* **43**, 83–98.
- MEI, Y. (2011). Quickest detection in censoring sensor networks. *2011 IEEE International Symposium on Information Theory Proceedings*, 2148–2152.
- MITCHELL, T. J. & BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **83**, 1023–1032.
- MONTGOMERY, D. C. (2020). *Introduction to statistical quality control*. New York: Wiley, 8th ed.
- NASA (2009). Solar dynamics observatory: Our eye on the sun. A guide to the mission and purpose of NASA’s Solar Dynamics Observatory. [https://www.nasa.gov/pdf/417176main\\_SDO\\_Guide\\_CMR.pdf](https://www.nasa.gov/pdf/417176main_SDO_Guide_CMR.pdf).
- OOMS, J. (2023). *av: Working with Audio and Video in R*. R package version 0.8.3.
- PAU, G., FUCHS, F., SKLYAR, O., BOUTROS, M. & HUBER, W. (2010). EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979–981.

- QIU, P. (2014). *Introduction to statistical process control*. Boca Raton: CRC Press.
- QIU, P. & HAWKINS, D. (2001). A rank-based multivariate CUSUM procedure. *Technometrics* **43**, 120–132.
- QIU, P. & HAWKINS, D. (2003). A nonparametric multivariate CUSUM procedure for detecting shifts in all directions. *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**, 151–164.
- R CORE TEAM (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society* **58**, 527–535.
- RUSSO, D. J., VAN ROY, B., KAZEROUNI, A., OSBAND, I. & WEN, Z. (2018). A tutorial on Thompson Sampling. *Foundations and Trends in Machine Learning* **11**, 1–96.
- SHEWHART, W. A. (1931). *Economic control of quality of manufactured product*. New York: D. Van Nostrand Company.
- TARTAKOVSKY, A., NIKIFOROV, I. & BASSEVILLE, M. (2014). *Sequential analysis: Hypothesis testing and changepoint detection*. New York: CRC press, 1st ed.
- THOMPSON, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25**, 285–294.
- WANG, A., XIAN, X., TSUNG, F. & LIU, K. (2018). A spatial-adaptive sampling procedure for online monitoring of big data streams. *Journal of Quality Technology* **50**, 329–343.
- WANG, Y. & BLEI, D. M. (2019). Frequentist consistency of variational bayes. *Journal of the American Statistical Association* **114**, 1147–1161.
- WANG, Y. & MEI, Y. (2013). Monitoring multiple data streams via shrinkage post-change estimation. *Submitted to Annals of Statistics* .

- XIAN, X., WANG, A. & LIU, K. (2018). A nonparametric adaptive sampling strategy for online monitoring of big data streams. *Technometrics* **60**, 14–25.
- XIAN, X., ZHANG, C., BONK, S. & LIU, K. (2021). Online monitoring of big data streams: A rank-based sampling algorithm by data augmentation. *Journal of Quality Technology* **53**, 135–153.
- XIE, Y., HUANG, J. & WILLETT, R. (2012). Change-point detection for high-dimensional time series with missing data. *IEEE Journal of Selected Topics in Signal Processing* **7**, 12–27.
- YAN, H., GRASSO, M., PAYNABAR, K. & COLOSIMO, B. M. (2022). Real-time detection of clustered events in video-imaging data with applications to additive manufacturing. *IIEE Transactions* **54**, 464–480.
- YAN, H., PAYNABAR, K. & SHI, J. (2015). Image-based process monitoring using low-rank tensor decomposition. *IEEE Transactions on Automation Science and Engineering* **12**, 216–227.
- YAN, H., PAYNABAR, K. & SHI, J. (2018). Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics* **60**, 181–197.
- YANG, W., GRASSO, M., COLOSIMO, B. & PAYNABAR, K. (2022). A tensor-based hierarchical process monitoring approach for anomaly detection in additive manufacturing. *Quality and Reliability Engineering International* **39**, 1–21.
- YE, H., XIAN, X., CHENG, J.-R. C., HABLE, B., SHANNON, R. W., ELYADERANI, M. K. & LIU, K. (2023). Online nonparametric monitoring of heterogeneous data streams with partial observations based on Thompson Sampling. *IIEE Transactions* **55**, 1–13.
- ZHANG, C. & HOI, S. (2019). Partially observable multi-sensor sequential change detection: A combinatorial multi-armed bandit approach. *Proceedings of the AAAI Conference on Artificial Intelligence* **33**, 5733–5740.

---

ZHANG, W. & MEI, Y. (2022). Bandit change-point detection for real-time monitoring high-dimensional data under sampling control. *Technometrics* **65**, 33–43.