

# La tecnica del video-monitoraggio

## *Procedure di pre-processamento*



Redatto da

Luca Parlagreco e Saverio Devoti

COS-CLM

In questa nota vengono descritte le caratteristiche e le modalità di funzionamento delle stazioni di video-monitoraggio utilizzate e le procedure pre-processamento utili per la successiva elaborazione.

**Le stazioni di video-monitoraggio ISPRA**

Id stazione	attività													
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
torresole01	*	*	*	*	*	*	*	*						
kufra														
TorreCerrano01														
senigallia														
rodi01														
fogliano01														

\* immagini prodotte da altro sistema di video-monitoraggio.

Il sistema di video-monitoraggio utilizzato è composto da due videocamere digitali orientate per riprendere una porzione di litorale di almeno 1km con una vista di quasi 180°.

Ogni videocamera è collegata ad un modulo di gestione che ne gestisce e ne sincronizza le attività. Tutti i dati video sono inviati sul server videomonitoraggio.isprambiente.it.

L’acquisizione delle foto è impostata in modo tale da garantire almeno cinque foto al giorno, negli orari che vanno dalle 11 alle 15 del pomeriggio, evitando i riverberi nocivi ai sensori che si generano per il riflesso del sole sul mare nei suoi momenti di salita e tramonto.

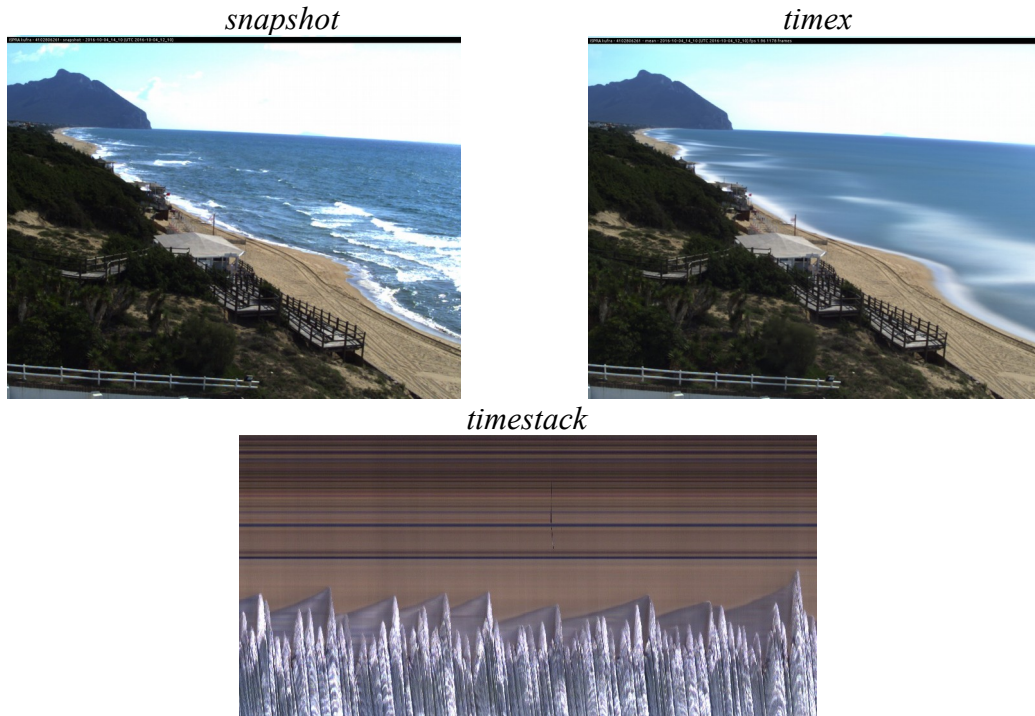
Le immagini fornite dalla stazione sono di tre tipi, *snapshot*, *timex* e *timestack* (figura 1).

La *snapshot* è una foto istantanea che documenta semplicemente l’attività del sistema spiaggia, ovvero ciò che accade al momento dall’acquisizione, è quindi utile ad esempio per osservare i ripascimenti, le caratteristiche del campo d’onda.

Le immagini con maggiore potenzialità nel fornire informazioni quantitative di tipo morfologico e idrodinamico sono la *time-exposure* (o *timex*) e la *timestack*, rispettivamente.

La *timex* è un’immagine in cui l’intensità luminosa dei pixel (RGB) è la media temporale di tutti i fotogrammi acquisiti con frequenza 2Hz in 10 minuti di registrazione. Da questo tipo di immagine si può sintetizzare la morfologia della spiaggia emersa e sommersa tramite la mappatura di alcuni video-elementi che sono, più o meno direttamente, connessi alla dinamica del moto ondoso. Ad esempio, la posizione della frangenza a riva e sulle barre sommerse. Come si vede da figura 2, la

posizione dei frangenti è ben documentata da bande chiare (alta luminosità) che si formano in conseguenza al processo di media temporale.

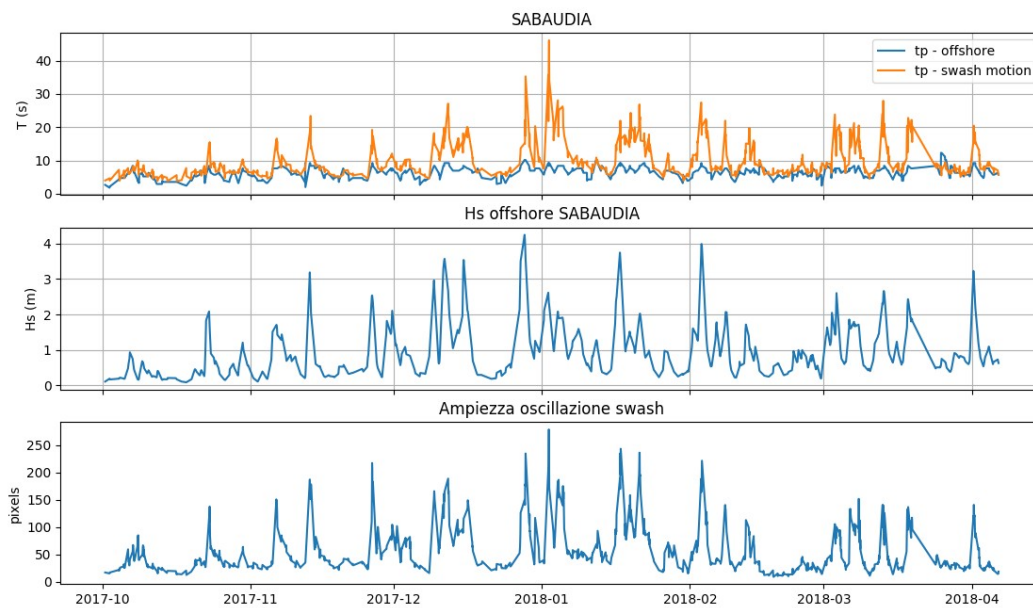


**Figura 1.** – Immagini *snap*, *timex* e *timestack* eseguite il giorno 2016-10-04 ore 14:10 dalla stazione di Kufra (Sabaudia).



**Figura 2.** – Immagine *timex* in cui sono ben visibili le due zone ad alta intensità luminosa indicative di settori preferenziali di frangimento ondoso (barre sommerse). Stazione di Rodi Garganico.

La *timestack* è un'immagine identificativa di una specifica porzione di fotogramma, generalmente un vettore di pixel tracciato perpendicolarmente alla linea di riva, per il quale si archivia l'intensità luminosa dei pixel (RGB) contestualmente alla costruzione della *timex*. Da questo tipo di immagine si ottengono indicazioni quantitative sull'oscillazione della risalita ondosa sulla spiaggia (*swash*) in termini di ampiezza e frequenza (figura 3).



**Figura 3.** – Serie temporale di ampiezza e frequenza di swash desunte da yimestack. Stazione di kufra (Sabaudia).

### Rilievi topografici a supporto del video-monitoraggio

Contestualmente all'installazione delle stazioni si eseguono una serie di rilievi mirati alla misura di particolari punti di controllo (*Ground Control Point, GCP*) utili sia alla procedura di rettifica delle foto, sia ad analizzare le accuratezze e gli errori intrinseci nell'elaborazione delle immagini video. Per ogni *GCP* misurato (figura 4) sono state archiviate le coordinate spaziali nel sistema di riferimento locale ( $x,y,z$ ) e le relative coordinate nel piano immagine.

In particolare sono state utilizzate delle mire graduate di 1m x 1m posizionate sulla spiaggia a diverse distanze dalla stazione di video-monitoraggio e utilizzate per stimare l'andamento delle accuratezze della rettifica in funzione della distanza dalla stazione stessa (figura 5).

Con un maggior grado di errore, sono state misurate una serie di boe di salvataggio (Agosto 2009) per analizzare gli stessi errori di rettifica nella zona delle barre.



**Figura 4.** – Posizionamento sul piano immagine dei GCPs utilizzati.



**Figura 5.** – Particolare della foto scattata in data 18 Febbraio 2009, in cui sono state posizionate delle mire graduate di 1m x 1m, evidenziate dalle frecce.

### **Procedure di image processing**

Un'immagine digitale è una rappresentazione bidimensionale discreta di una porzione tridimensionale dello spazio continuo.

Tale rappresentazione avviene per mezzo di determinati processi di campionamento, in cui la misura della luce è demandata a sensori elettronici che convertono l'intensità del segnale luminoso in un impulso elettrico semplice. Le pellicole, diversamente, traducono l'intensità della luce ed il colore tramite processi chimico-fisici della pellicola fotografica. La rappresentazione dello spazio fotografato è detta 'discreta' in quanto la foto digitale è composta da un numero finito di elementi principali, detti *pixels*.

Generalmente, un'immagine RGB codifica, per ogni *pixel*, il colore per mezzo di tre livelli o bande di colore (R rosso; G verde, B blu). Ne risulta che, un'immagine digitale  $a[m,n,i]$ , sarà costituita da  $m \times n$  *pixels*, di cui  $m$ -*pixels* lungo l'asse verticale della foto ed  $n$ -*pixels* lungo quello orizzontale. Ulteriormente gli  $m \times n$  *pixels* della foto sono ripetuti per ogni spazio di colore  $i$ , nel caso RGB  $i=3$  ( $i=1$  rosso,  $i=2$  verde,  $i=3$  blu). In termini puramente matematici la foto digitale è una matrice dove ogni *pixel* è caratterizzato da una coppia di coordinate (coordinate immagine), di seguito chiamate  $u$  e  $v$  ( $u$  = riga e  $v$  = colonna). Ogni processo applicato ad una foto digitale risulta quindi da una serie di operazioni matematiche eseguite per ogni elemento (*pixel*) della matrice (foto).

Il video-monitoraggio utilizza come dato di *input* una visione prospettica di un determinato tratto litorale (fotografia). Vengono utilizzate, quindi, una serie di procedure matematiche per eliminare l'influenza della distorsione prospettica sulla rappresentatività spaziale della foto, e quindi per analizzare la foto in chiave metrica (fotogrammetria) ed estrarre informazioni circa lunghezze e relazioni tra lunghezze desumibili da foto.

Il grado di accuratezza delle potenziali informazioni spaziali estraibili da una foto sono funzione di:

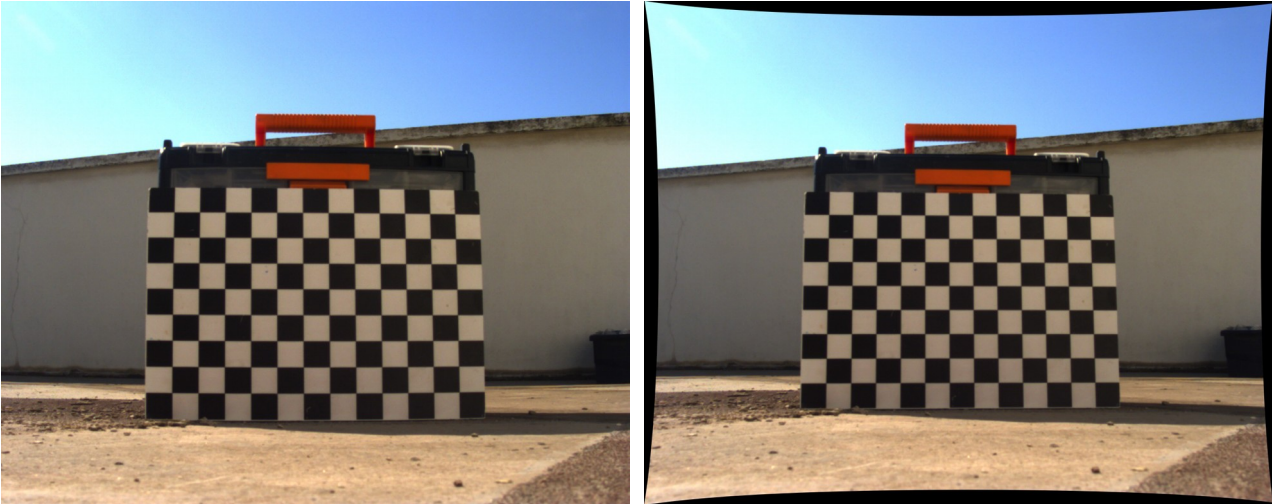
- orientazione spaziale della camera (definita dalla sua geometria esterna);
- caratteristiche intrinseche del sistema di ripresa (definite dalla sua geometria interna).

Mentre definire i parametri della geometria esterna è funzionale alla geo-referenziazione del dato immagine, conoscere i parametri intrinseci della foto-camera utilizzata (geometria interna) è funzionale alla correzione delle distorsioni che gli obiettivi sferici impartiscono all'immagine fotografata, e quindi alla rappresentatività spaziale della foto. In figura 6 si riporta il fotogramma originale e lo stesso a seguito di rimozione della distorsione delle ottiche.

La calibrazione di una camera digitale rappresenta una specifica procedura necessaria per definire, in maniera quanto più esatta possibile, i parametri caratteristici della geometria interna ed esterna di un sistema di ripresa. Una volta determinati tali parametri, è possibile calcolare la relazione geometrica tra coordinate immagine e coordinate spaziali, ovvero è possibile proiettare una foto digitale prospettica  $a[m,n,i]$  in un nuovo sistema di coordinate piane in cui l'unità di misura dei *pixels* può rappresentare un'unità di misura metrica, la nuova foto digitale  $r[p,q,i]$  è detta rettificata (o ribaltata).

La risoluzione spaziale della foto  $r[p,q,i]$  sarà ben differente, quindi, dalla risoluzione della foto originale  $a[m,n,i]$ , è sarà funzione della maglia di ri-campionamento prescelta durante il calcolo della proiezione dal piano immagine al piano 'spiaggia' (processo di rettifica). Ad esempio se  $p=1000$  e  $q=500$ , l'immagine  $r$  è interpretabile come una mappa topografica di una porzione di

territorio di 500.000 unità. La porzione sarà di 500.000mq se ogni *pixel* dell'immagine rettificata rappresentasse una porzione di territorio di  $1m \times 1m$ .



**Figura 6.** – Immagine *timex* non calibrata (sx) e calibrata (dx) da distorsione delle ottiche.

### Calcolo della geometria interna

I principali parametri che caratterizzano la geometria interna di un sistema di ripresa sono:

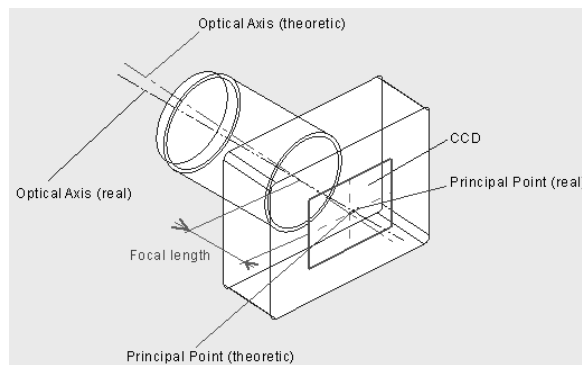
- a. posizione del punto principale all'interno del piano immagine (*principal point*);
- b. focale effettiva (*focal length*);
- c. grado di asimmetria dei *pixel* (*skew coefficient*);
- d. coefficienti di distorsione ottica delle lenti (*distorsion coefficients*).

In figura 7 vengono raffigurati i parametri *a* e *b*. Per le camere non-metriche (quali quelle utilizzate dal sistema ISPRA) il posizionamento del punto principale effettivo non coincide necessariamente con il centro geometrico della foto. Analogamente, la lunghezza della focale effettiva non necessariamente coincide con il valore nominale fornito dal costruttore.

La deviazione tra valori nominali e valori effettivi dei parametri in oggetto può pregiudicare notevolmente il rendimento di un sistema di ripresa, per tale motivo la calibrazione della camera viene utilizzata per ottimizzare il rendimento del sistema, ovvero per raffinare la procedura di calcolo che cerca la relazione tra sistema di riferimento immagine e sistema di riferimento spaziale.

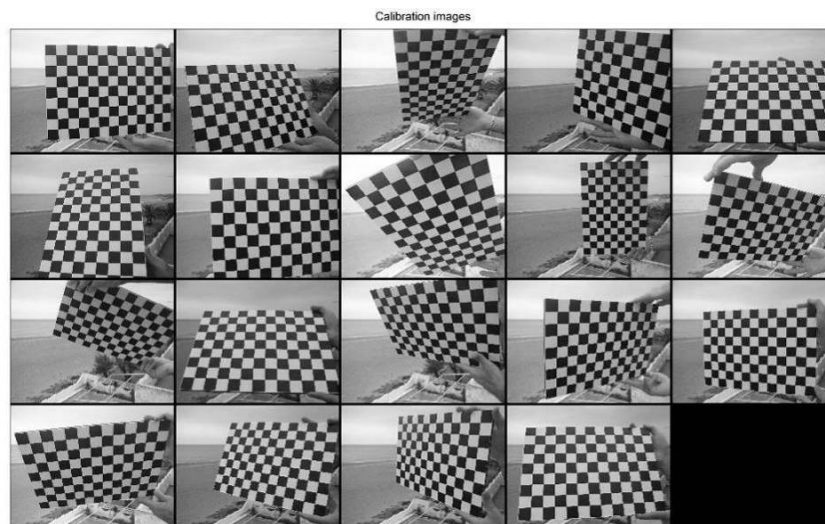
La calibrazione di un sistema di ripresa, soprattutto, fornisce un modo per conoscere e correggere l'errore dovuto alla distorsione ottica dell'obiettivo.

A tale scopo, si è utilizzato un toolbox<sup>1</sup> sviluppato in ambiente Matlab che permette di calcolare gli errori relativi alla deformazione delle ottiche e quindi a rettificare le immagini. Il metodo utilizzato è riconducibile al metodo di Zhang (2000). Questo metodo si basa sul *pinhole camera model*, ovvero si basa sul principio che la stazione di ripresa si fissa nello spazio, come nel sistema di video-monitoraggio.



**Figura 7** - Struttura schematica di una camera digitale.

L'approccio utilizza una serie di regole della geometria proiettiva per calcolare, iterativamente, e indirettamente le caratteristiche interne della macchina. La procedura si basa sull'utilizzo di una mira, generalmente costituita da una scacchiera piana, fotografata in diverse orientazioni rispetto l'orientazione della foto-camera da calibrare (figura 8).

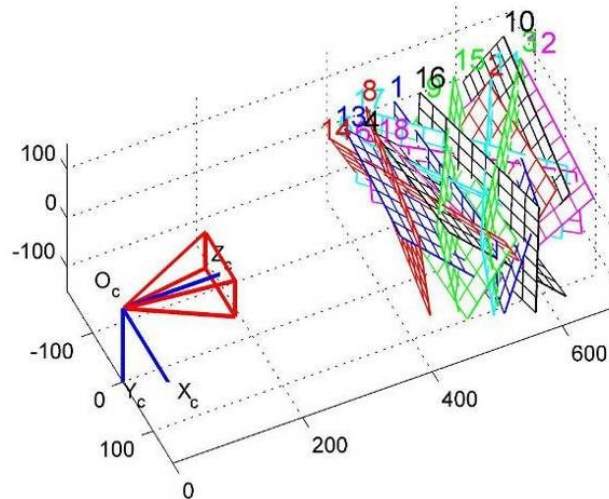


**Figura 8.** - Immagini utilizzate per la calibrazione di una foto-camera della stazione *Torre del Sole*.

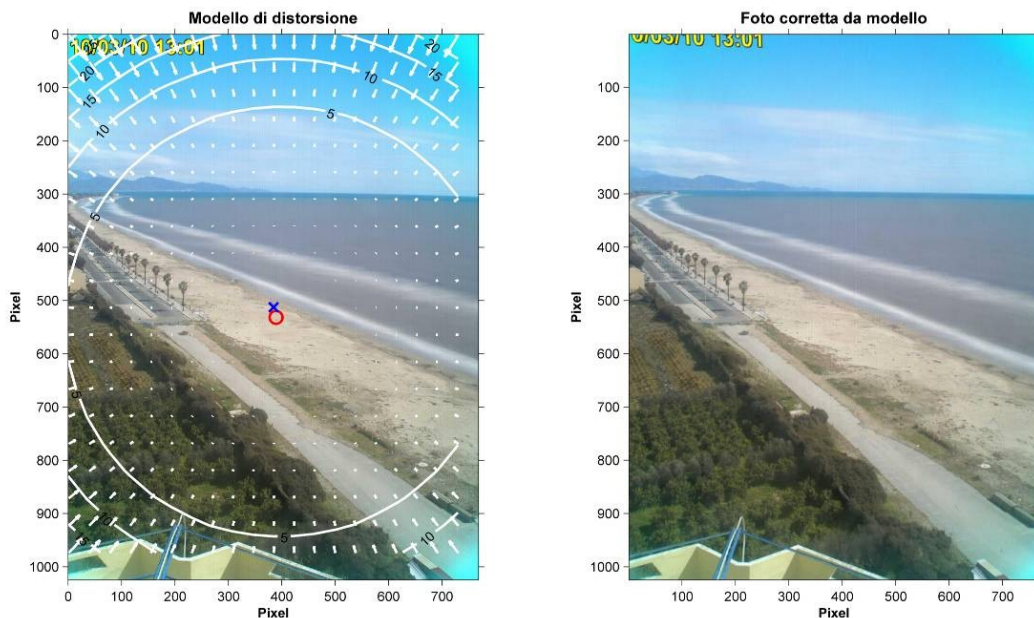
Caricate le immagini della scacchiera nel toolbox la procedura di calibrazione si effettua selezionando, per ogni istantanea, un'area della scacchiera in cui i relativi riquadri siano ben visibili e nitidi. L'algoritmo sviluppato da Zhang estrae automaticamente i vertici della griglia fotografata

<sup>1</sup> Camera Calibration Toolbox for Matlab®, disponibile al [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)

per ognuna delle orientazioni, al fine di computare una trasformazione proiettiva tra i punti immagine delle  $n$  immagini scattate (di solito 15/20 foto). Conoscendo le dimensioni reali dei riquadri, calcola tutti i parametri caratteristici della camera con i relativi errori. Successivamente, per ogni orientazione della scacchiera è stato possibile stabilire la relazione spaziale tra il posizionamento della videocamera e l'orientazione del piano fotografato (figura 9).



**Figura 9.** - Orientazione delle scacchiere fotografate relativa alla videocamera del sistema *Torre del Sole*.  $O_c$  rappresenta la posizione della foto-camera in coordinate locali.



**Figura 10.** – Immagine originale ed errori di distorsione delle ottiche (modello di distorsione calcolato) e immagine corretta da modello. Al centro della foto di sinistra sono riportati i Punti Principali nominali ed effettivi. Si noti, inoltre la distorsione della scritta in alto a sinistra nella foto di destra a testimonianza della correzione della distorsione ottica sull'immagine fotografica.

Una volta definita la relazione tra mire fotografate e videocamera, conoscendo anche la spaziatura della griglia utilizzata, è possibile risolvere i parametri interni della videocamera ai minimi quadrati. Ulteriormente, tale metodo minimizza gli errori di proiezione e raffina l'estrazione dei parametri interni. L'errore di proiezione indica la risoluzione con cui il modello calcola i parametri interni della macchina, e quindi le distorsioni delle ottiche utilizzate (figura 10).

Un ulteriore parametro utile alla calibrazione del sistema è il *field of view (FOV)*, indicante l'apertura angolare effettiva della foto-camera in funzione della lunghezza focale utilizzata.

### Calcolo della geometria esterna

I parametri di orientamento esterno definiscono la posizione della foto-camera nello spazio reale, il loro calcolo permette di stabilire la relazione geometrica tra le coordinate immagine ( $u, v$ ) e le coordinate spaziali ( $x, y, z$ ).

I parametri che definiscono la geometria esterna di un sistema di ripresa sono definiti da tre valori angolari (figura 11): i) l'azimuth ( $\varphi$ ); ii) il tilt ( $\tau$ ); iii) il roll ( $\sigma$ ).

La relazione tra coordinate immagine ( $u, v$ ) a coordinate spaziali ( $x, y, z$ ) è formalizzata a mezzo di una serie di coefficienti lineari:

$$u = \frac{L_1 x L_2 y L_3 z L_4}{L_9 x L_{10} y L_{11} z 1} \quad e \quad v = \frac{L_5 x L_6 y L_7 z L_8}{L_9 x L_{10} y L_{11} z 1} \quad (3.1)$$

dove,  $L_1-L_{11}$  rappresentano funzioni lineari di sette incognite:  $\tau, \varphi, \sigma, [x_c, y_c, z_c]$  (posizione del sistema di ripresa) ed  $f$  (focale effettiva):

$$\begin{aligned} L &= -(x_c m_{31} + y_c m_{32} + z_c m_{33}) \\ L_1 &= \frac{(u_0 m_{31} + f m_{11})}{(\lambda_u L)} \\ L_2 &= \frac{(u_0 m_{32} + f m_{12})}{(\lambda_u L)} \\ L_3 &= \frac{(u_0 m_{33} + f m_{13})}{(\lambda_u L)} \\ L_4 &= -(L_1 x_c + L_2 y_c + L_3 z_c) \\ L_5 &= \frac{(v_0 m_{31} + f m_{21})}{\lambda_v L} \\ L_6 &= \frac{(v_0 m_{32} + f m_{22})}{\lambda_v L} \\ L_7 &= \frac{(v_0 m_{33} + f m_{23})}{\lambda_v L} \\ L_8 &= -(L_8 x_c + L_9 y_c + L_{10} z_c) \\ L_9 &= \frac{m_{31}}{L} \\ L_{10} &= \frac{m_{32}}{L} \\ L_{11} &= \frac{m_{33}}{L} \end{aligned} \quad \begin{aligned} L &= -(x_c m_{31} + y_c m_{32} + z_c m_{33}) \\ L_1 &= \frac{(u_0 m_{31} + f m_{11})}{(\lambda_u L)} \\ L_2 &= \frac{(u_0 m_{32} + f m_{12})}{(\lambda_u L)} \\ L_3 &= \frac{(u_0 m_{33} + f m_{13})}{(\lambda_u L)} \\ L_4 &= -(L_1 x_c + L_2 y_c + L_3 z_c) \\ L_5 &= \frac{(v_0 m_{31} + f m_{21})}{\lambda_v L} \\ L_6 &= \frac{(v_0 m_{32} + f m_{22})}{\lambda_v L} \\ L_7 &= \frac{(v_0 m_{33} + f m_{23})}{\lambda_v L} \\ L_8 &= -(L_8 x_c + L_9 y_c + L_{10} z_c) \\ L_9 &= \frac{m_{31}}{L} \\ L_{10} &= \frac{m_{32}}{L} \\ L_{11} &= \frac{m_{33}}{L} \end{aligned}$$

Gli  $m$ -coefficienti descrivono le successive rotazioni rispetto l'azimuth ( $\varphi$ ), il tilt ( $\tau$ ) e il roll ( $\sigma$ ):

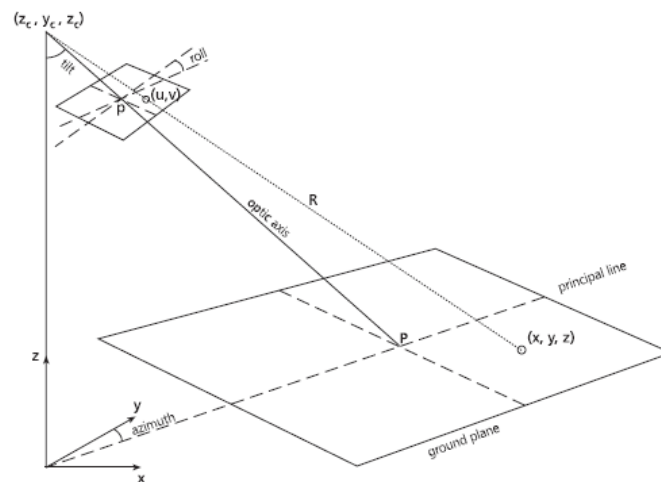
$$\begin{aligned} m_{11} &= \cos \phi \cos \sigma + \sin \phi \cos \tau \sin \sigma \\ m_{12} &= -\sin \phi \cos \sigma + \cos \phi \cos \tau \sin \sigma \\ m_{13} &= \sin \tau \sin \sigma \\ m_{21} &= -\cos \phi \sin \sigma + \sin \phi \cos \tau \cos \sigma \\ m_{22} &= \sin \phi \sin \sigma + \cos \phi \cos \tau \cos \sigma \\ m_{23} &= \sin \tau \cos \sigma \\ m_{31} &= \sin \phi \sin \tau \\ m_{32} &= \cos \phi \sin \tau \\ m_{33} &= -\cos \tau \end{aligned} \quad \begin{aligned} m_{11} &= \cos \phi \cos \sigma + \sin \phi \cos \tau \sin \sigma \\ m_{12} &= -\sin \phi \cos \sigma + \cos \phi \cos \tau \sin \sigma \\ m_{13} &= \sin \tau \sin \sigma \\ m_{21} &= -\cos \phi \sin \sigma + \sin \phi \cos \tau \cos \sigma \\ m_{22} &= \sin \phi \sin \sigma + \cos \phi \cos \tau \cos \sigma \\ m_{23} &= \sin \tau \cos \sigma \\ m_{31} &= \sin \phi \sin \tau \\ m_{32} &= \cos \phi \sin \tau \\ m_{33} &= -\cos \tau \end{aligned}$$

Queste incognite possono essere calcolate tramite l'ausilio di una serie di punti di riferimento nell'immagine (*GCP*) di cui siano note sia le coordinate immagine ( $u, v$ ) sia le coordinate spaziali ( $x, y, z$ ).

Sebbene questa procedura riesca a calcolare indirettamente il posizionamento della foto-camera ( $x_c, y_c, z_c$ ), Holland et al. (1997) indicano che la loro misura diretta fornisce una maggiore accuratezza della trasformazione. Per tale motivo, sono state misurate le coordinate spaziali del baricentro dei sistemi di ripresa.

Conoscendo almeno tre *GCPs* è possibile risolvere le equazioni di collinearità e calcolare le incognite (parametri della geometria esterna,  $\tau, \varphi, \sigma$ ). Il calcolo del *FOV* ( $\delta$ ) è stato ottenuto mediante la procedura di calibrazione del modello.

In seguito ad una serie di prove, utilizzando un totale di circa 30 *GCPs* misurati sulla spiaggia si è optato per l'utilizzo di 5 *GCPs* per la camera di levante (S3) e 4 *GCPs* per la camera di ponente (S4).



**Figura 11.** - Relazione tra coordinate immagine ( $u, v$ ) e coordinate reali ( $x, y, z$ ).

### Geo-referenziazione e ribaltamento delle foto

La geo-referenziazione delle immagini è funzione dei parametri di geometria esterna della foto-camera ( $\tau, \varphi, \sigma, x_c, y_c, z_c$ ) e dalla focale effettiva  $f$ .

Partendo dalla seguente equazione è possibile geo-referire le immagini:

$$\begin{bmatrix} L_1 - L_9u & L_2 - L_{10}u & L_3 - L_{11}u \\ L_5 - L_9v & L_6 - L_{10}v & L_7 - L_{11}v \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} v - L_4 \\ v - L_8 \end{bmatrix}. \quad (1)$$

dove i coefficienti  $L-L_{11}$  sono stati precedentemente descritti. La trasformazione di ogni *pixel* dell'immagine obliqua  $(u,v)$  permette di localizzare nello spazio  $(x,y,z)$  una determinata porzione di spiaggia fotografata. Dal punto di vista matematico, questa trasformazione è di tipo bidimensionale, ovvero trasforma il piano immagine (foto obliqua) nel piano spiaggia, assumendo una quota di riferimento ( $z$ ) costante, solitamente il livello di marea al momento di scatto. Generalmente, questo significa che le coordinate  $xy$  vengono proiettate su un piano la cui quota è coincidente, per esempio, al livello mareale. Quando la quota di rettifica  $z$  utilizzata non è ben rappresentativa delle condizioni idrodinamiche a costa si possono introdurre degli errori nel calcolo delle coordinate  $xy$  che complicheranno l'interpretazione morfologica della foto (si veda in seguito).

Nel nostro caso, la su menzionata relazione è stata calcolata per ogni *pixel* della foto prospettica (corretta dalla distorsione ottica) e ha fornito una nuova foto ribaltata caratterizzata da una risoluzione di 1m x 1m per il tratto di litorale posizionato tra 0m e -1500m lungo costa e da 0m a 700m trasversalmente a costa.

#### **Accuratezze e risoluzione spaziale nominale delle foto ribaltate**

Le immagini ribaltate e rettificata costituiscono la base per tutte le successive analisi e descrizioni dei dati video-morfologici. L'accuratezza di questi dati limita a sua volta l'accuratezza di tutti i parametri morfologici estratti da foto. Le accuratezze della rettifica dipendono dalle accuratezze della calibrazione delle foto-camere e dalla quota di riferimento ( $z$ ) del piano di proiezione. L'accuratezza della calibrazione viene espressa da un errore spaziale della foto. Esso rappresenta la distanza euclidea tra un punto misurato (visibile nella foto) e la sua posizione nel piano fotografico calcolata a partire dalle coordinate  $xyz$  tramite l'equazione 1. Questa procedura è stata eseguita per ogni GCPs proprio per tarare la risoluzione geometrica più appropriata. Gli errori spaziali calcolati risultano generalmente inferiori al *pixel*.

La prospettiva induce una capacità di risoluzione spaziale della foto che varia in funzione dell'angolo che l'asse ottico forma con la superficie che si vuole fotografare. Generalmente si definisce *pixel footprint* la quantità di spazio che viene rappresentato da un singolo *pixel* della foto. Va da se che in una visione prospettica, quale una presa fotografica, il *pixel footprint* varierà sistematicamente in funzione della distanza dal punto di scatto.

Generalmente il *pixel footprint* rappresenta una porzione rettangolare di territorio fotografato, in cui il lato maggiore  $dy$  (in foto effettuate con presa orizzontale) è funzione della distanza dalla camera di scatto ( $R$  in figura 11) e del  $FOV$   $\delta$  (in radianti), secondo la relazione

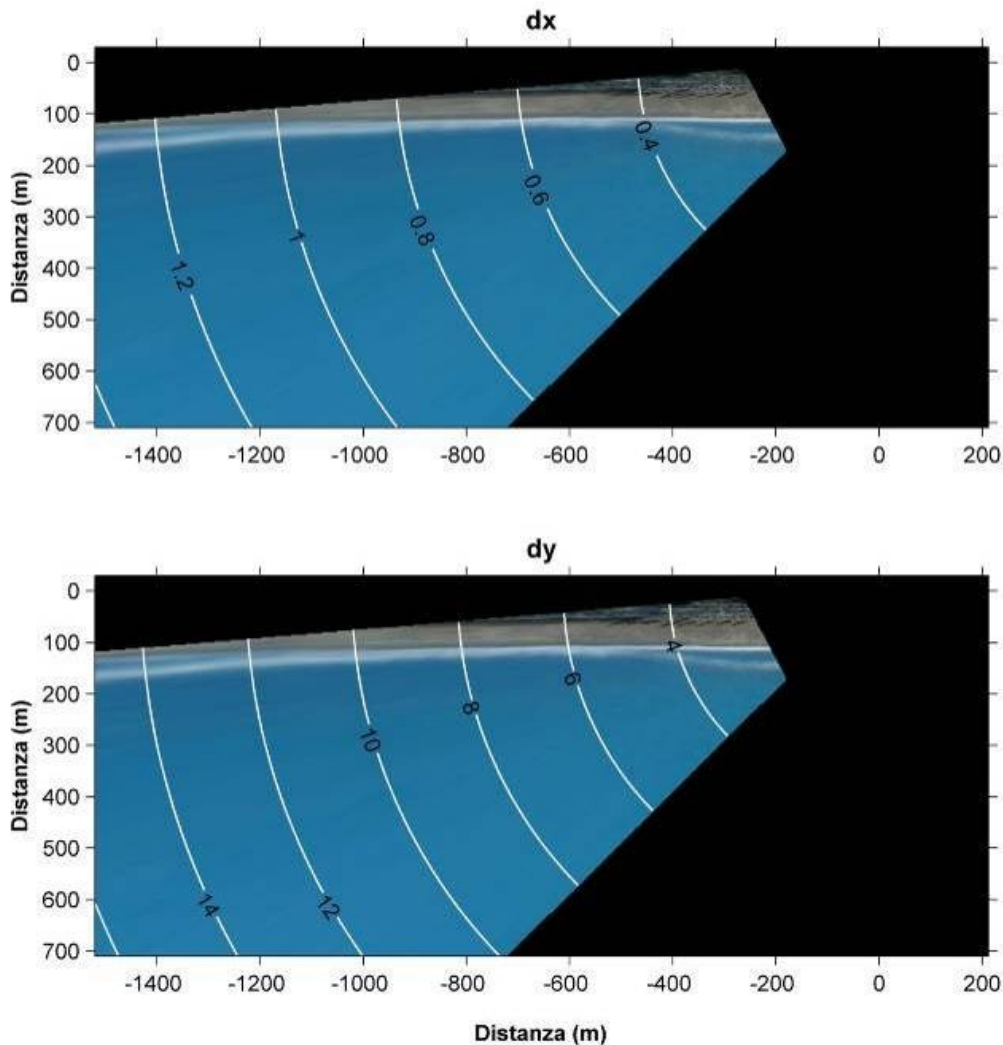
$$dy = R \delta / \text{numero di pixel orizzontali} \quad (2)$$

Il lato minore  $dx$  del *pixel footprint* dipende dall'angolo di tilt  $\tau$ , e viene calcolato come,

$$dx = \tan(\tau) R \delta / \text{numero di pixel orizzontali} \quad (3)$$

Il *pixel footprint* è spesso orientato obliquamente rispetto alla linea di riva, in quanto le prese fotografiche di video-monitoraggio hanno una visione obliqua della spiaggia. Per tale motivo, sia  $w$  che  $l$  deviano leggermente dalla risoluzione *cross-shore* e *alongshore*. Utilizzando l'angolo tra l'azimuth e l'orientazione della linea di riva,  $dx$  e  $dy$  possono essere trasformati in risoluzioni *cross-shore* e *alongshore*. Figura 12 mostra queste dimensioni calcolate per la stazione *Torre del Sole* (setting 2009-2014).

Le dimensioni  $dx$  del *pixel footprint* variano in tutta l'area di video-monitoraggio da 0.4m a 1.2m, ben inferiori della maglia di ri-campionamento utilizzata nella procedura di rettifica (2.5m). Le dimensioni lungo costa ( $dy$ ) dello stesso variano tra i 4m in vicinanza della camera ( $x = -400$ ) a 14m a circa 1500m dalla camera.



**Figura 12.** – Andamento del *pixel footprint* (in m) per la stazione Torre del Sole (setting 2009-2014).