

## L'OPERATORE SIFT PER L'ORIENTAMENTO DI IMMAGINI ACQUISITE CON PRESE NON-NORMALI

Andrea LINGUA, Davide MARENCHINO, Francesco NEX

(\*) DITAG – Dipartimento di Ingegneria del Territorio, dell'Ambiente e delle Geotecnologie, Politecnico di Torino  
C.so Duca degli Abruzzi 24, 10129 Torino – (andrea.lingua, davide.marenchino, francesco.nex)@polito.it

### Riassunto

In questi ultimi anni lo sviluppo di strumenti fotogrammetrici in ambito aereo si è evoluto fortemente. Allo stato attuale le piattaforme aeree autopilotate (UAV) e i sistemi per l'acquisizione di immagini oblique forniscono immagini con geometrie di presa lontane dalla condizione di normalità. In questo contesto, un'alternativa agli operatori di interesse per la *feature extraction* utilizzati in campo aereo (Forstner, Harris, ecc.) è costituita dagli algoritmi proposti in computer vision e in fotogrammetria dei vicini. In questo articolo viene testato l'operatore SIFT (Lowe, 2004) su immagini terrestri e su immagini aeree per valutarne le potenzialità applicative.

### Abstract

*The development of photogrammetric aerial systems has changed the user demand. Nowadays the increasing use of unmanned aerial vehicles (UAV) and new oblique aerial cameras acquire images far from normal conditions. For this reason traditional interest operators (Forstner, Harris, etc.) are joined by new algorithms already proposed in computer vision. In this paper, some tests of the SIFT operator (Lowe, 2004) are performed considering both terrestrial and aerial images, in order to evaluate the performance of this operator.*

### INTRODUZIONE

L'operazione di *feature extraction* è un passo fondamentale nel processamento di immagini per applicazioni fotogrammetriche e di computer vision. Da questa fase dell'elaborazione dipendono le fasi di orientamento automatico e semiautomatico delle immagini, e la definizione dei *seed point* del *matching* denso. Sono allo studio numerosi operatori di interesse in grado di estrarre e accoppiare punti significativi in condizioni critiche; attualmente l'algoritmo SIFT (*Scale-Invariant Feature Transform*) è uno degli operatori sui quali si è maggiormente rivolta l'attenzione del mondo della ricerca (Mikolajczyk, Schmid, 2005). L'algoritmo è stato implementato dagli autori in ambiente Matlab, allo scopo di valutarne le prestazioni in vari ambiti applicativi. L'operatore SIFT è stato quindi testato su immagini aeree e terrestri ed i risultati ottenuti sono stati confrontati con quelli forniti dal ben noto operatore di Forstner.

### L'operatore SIFT

L'operatore SIFT è un algoritmo in grado di estrarre (*detector*) e di "descrivere" (*descriptor*) *features* su immagini stereoscopiche per la ricerca di punti omologhi. Gli operatori utilizzati tradizionalmente in fotogrammetria (Forstner, Gulch 1988; Harris, Stephens, 1988) ricercano punti omologhi (*point detector*) su spigoli o discontinuità radiometriche; al contrario l'operatore SIFT ricerca i punti (*keypoint*) su regioni più ampie dell'immagine (*region detector*) superando i problemi di occlusione e deformazione prospettiche (Lowe, 2004). A tale scopo, ogni immagine è innanzitutto filtrata mediante convoluzioni gaussiane, formando uno spazio delle "scale". Per ogni

scala sono quindi calcolate le differenze fra gaussiane adiacenti (*DoG*, *Difference of Gaussian*) i cui massimi sono memorizzati come punti di interesse (*keypoints*). La ricerca di questi punti nello spazio delle scale anziché sulle immagini originali fa sì che i *keypoints* estratti non coincidano con spigoli o discontinuità radiometriche ma si localizzino in prossimità degli stessi. Per ogni *keypoint* individuato è quindi definito un descrittore (vettore di numeri) capace di descrivere i gradienti radiometrici nell'intorno del punto di interesse indipendentemente da rotazioni, variazioni di scala e cambiamenti di illuminazione. In base alla distanza euclidea fra questi vettori  $n$ -dimensionali è infine possibile individuare *keypoints* omologhi fra le immagini. L'algoritmo prevede la definizione di alcuni parametri da cui, come sarà proposto nel seguito, dipende il numero di punti estratti e il tempo di elaborazione per ciascuna immagine. Per una descrizione completa dell'operatore si rimanda comunque a [Lowe, 2004].

### Eliminazione degli outlier

La principale caratteristica che distingue l'operatore SIFT dagli altri operatori di interesse utilizzati per il *features extraction* è la robustezza. Le invarianze alla scala, alla rotazione e all'illuminazione garantiscono l'estrazione e l'individuazione automatica di punti omologhi anche in condizioni "lontane" dall'ipotesi di presa normale. In questo caso l'aggettivo "robusto" è esclusivamente legato alla capacità di estrarre e associare punti di interesse. L'operazione di *matching* proposta da Lowe (Lowe, 2004) non garantisce però robustezza nella definizione delle coppie omologhe. Permangono quindi i problemi legati a possibili accoppiamenti di punti omologhi errati, con il conseguente insorgere di errori grossolani (*outlier*).

Perciò, prima di procedere alla fase di *bundle block adjustment*, è necessario ricorrere a opportune tecniche di stima robusta per l'individuazione degli *outlier*. In particolare è stato utilizzato lo stimatore robusto LMS (*Least Median Square*, Rousseau, 1987) per il calcolo dell'orientamento relativo tra coppie di immagini omologhe.

## IMPLEMENTAZIONE E VALIDAZIONE

Al fine di validare l'algoritmo di calcolo (implementato in ambiente Matlab), e di testare le potenzialità dell'operatore SIFT, sono state svolte numerose prove su immagini di differente natura. Le analisi sono state compiute mediante il confronto con i risultati di *corner extraction*, forniti dall'operatore di Forstner, e di accoppiamento ottenuti con la tecnica di *cross correlation* (ForCC). L'estrazione di punti di interesse è avvenuta su immagini pretrattate con il filtro di *Wallis*. Tutti gli algoritmi di calcolo sono stati sviluppati in ambiente Matlab.

### Test su immagini sintetiche

Il primo test è stato eseguito su una coppia di immagini sintetiche (Remondino, 2006), che presentano una variazione di scala (fattore 0.9) ed una rotazione di circa 20°. La Figura 1 mostra i risultati del match. In verde sono evidenziate le corrispondenze corrette ed in rosso quelle errate.

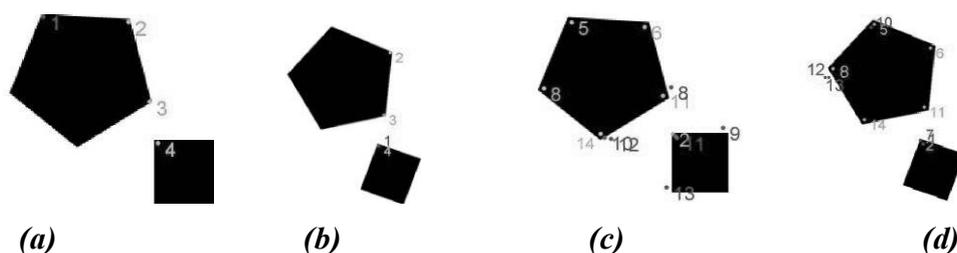


Figura 1 - Risultati di image matching utilizzando l'operatore di Forstner (a-b) e SIFT (c-d)

I risultati sperimentali (Tabella 2) confermano l'insensibilità di SIFT a rotazioni e variazioni di scala. Come mostrato in Figura 1, SIFT non estrae con accuratezza gli angoli, bensì dei punti in prossimità degli stessi. Si può infine notare che la percentuale di punti estratti correttamente è maggiore in ForCC (75% con soglia di correlazione pari a 0.8), anche se il numero complessivo di *corners* individuati è maggiore in SIFT. L'analisi delle corrispondenze è stata compiuta mediante confronto visivo.

	FORSTNER+CC	SIFT
<b>Punti omologhi estratti</b>	<b>4</b>	<b>14</b>
<b>Punti omologhi corretti</b>	<b>3</b>	<b>6</b>
% match corretti	75%	43%
<b>Corner estratti</b>	<b>3</b>	<b>6</b>
<b>N° totale corner</b>	<b>9</b>	<b>9</b>
% corner estratti	33%	67%

**Tabella 2 - analisi delle prestazioni degli operatori sulle immagini sintetiche**

### Test su immagini aeree

Un secondo test di validazione è stato effettuato su una coppia di immagini aeree di una zona boschiva, riprese con la camera *Intergraph's Z/I Imaging DMC*. Per ridurre i tempi di calcolo automatico, le immagini sono state ricampionate a 600 *dpi*. Lo scopo del test è stato quello di analizzare le percentuali di punti omologhi estratti in funzione delle soglie caratteristiche di ogni operatore e dello stimatore LMS. In particolare sono state effettuate differenti prove di estrazione e di matching variando il limite minimo del coefficiente di correlazione (in ForCC), il massimo rapporto di distanza euclidea tra i descrittori dei keypoint estratti  $th\_eu$  (in SIFT) per la ricerca dei punti omologhi e la soglia di reiezione  $L$  (in LMS).

I risultati sono riassunti in Tabella 3. Con  $nth$  si intende il numero di coppie omologhe estratte dagli operatori, e con  $nthok$  le coppie residue in seguito alla stima LMS.

$L$	FORSTNER+CC					SIFT				
		$nth$	$nthok$	%	$p_{max}$ (pixel)	$Th\_eu$	$nth$	$nthok$	%	$p_{max}$ (pixel)
3	<b>0.8</b>	134	35	26.1	0.074	<b>1.25</b>	2454	22	0.9	0.003
	<b>0.85</b>	92	25	27.1	0.089	<b>1.5</b>	2122	26	1.2	0.007
	<b>0.9</b>	46	27	<b>58.6</b>	<b>0.165</b>	<b>2</b>	1572	25	<b>1.6</b>	<b>0.006</b>
5	<b>0.8</b>	134	67	50	0.225	<b>1.25</b>	2454	76	3.1	0.013
	<b>0.85</b>	92	55	59.7	0.277	<b>1.5</b>	2122	81	3.8	0.014
	<b>0.9</b>	46	35	<b>76.0</b>	<b>0.413</b>	<b>2</b>	1572	66	<b>4.2</b>	<b>0.014</b>
10	<b>0.8</b>	134	123	91.7	0.823	<b>1.25</b>	2454	309	12.6	0.056
	<b>0.85</b>	92	91	98.9	1.071	<b>1.5</b>	2122	319	15.0	0.062
	<b>0.9</b>	46	46	<b>100</b>	<b>0.553</b>	<b>2</b>	1572	280	<b>17.8</b>	<b>0.072</b>
15	<b>0.8</b>	134	129	96.2	1.091	<b>1.25</b>	2454	616	25.1	0.122
	<b>0.85</b>	92	91	98.3	1.073	<b>1.5</b>	2122	651	30.7	0.148
	<b>0.9</b>	46	46	<b>100</b>	<b>0.553</b>	<b>2</b>	1572	677	<b>43.1</b>	<b>0.182</b>

**Tabella 3 - Risultati forniti dagli operatori in esame in funzione dei parametri  $th\_eu$ ,  $L$**

Con soglie di reiezione basse il numero di corrispondenze estratte dai due operatori è simile, ma analizzando le parallassi residue massime si può affermare che l'accuratezza di SIFT è decisamente superiore rispetto a ForCC. Al crescere del numero di reiezione il numero di corrispondenze mantenute dall'algorithm LMS aumenta, ciò a discapito della precisione. E' da notare che in ForCC alte soglie di correlazione permettono di estrarre coppie con alta probabilità di match corretto (100% con  $L=10$ ;  $\rho = 0.9$ ), mentre l'incremento della soglia di matching di SIFT non comporta una grande variazione nella percentuale di punti di match corretti. Si può comunque affermare che per alte soglie di reiezione il numero assoluto di coppie omologhe depurate da errori grossolani è decisamente superiore con l'uso di SIFT, e le parallassi residue sono comunque inferiori a quelle fornite da ForCC. Nelle prove eseguite i parametri di orientamento relativo sono stati confrontati con i parametri stimati mediante restitutore analitico e si mantengono sostanzialmente costanti (test t-Student).

## APPLICAZIONI FOTOGRAMMETRICHE DI SIFT

Attualmente nei campi della *computer vision* (*motion tracking*, *3D reconstruction*) e della fotogrammetria dei vicini, la geometria di acquisizione delle immagini non è più vincolata alla condizione di presa "normale". Lo stesso concetto vale per particolari applicazioni aeree, quali l'utilizzo di *mini-UAV* (*Unmanned Aerial Vehicle*, UAV association, 2007) per l'acquisizione stereoscopica di immagini e i sistemi per l'acquisizione di immagini oblique. In questi casi l'operatore SIFT è uno dei principali candidati per le operazioni di *feature extraction*. A tal proposito sono stati compiuti test su immagini per applicazioni terrestri ed aeree.

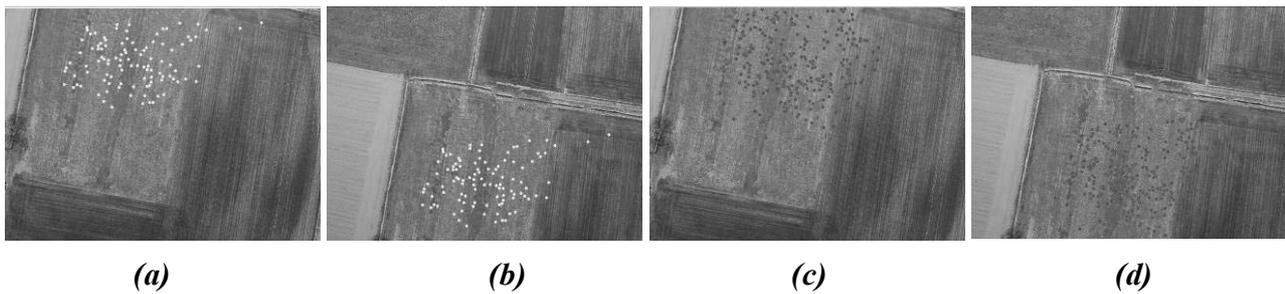
### Applicazioni aeree (UAV)

L'uso di UAV per scopi fotogrammetrici è in continua crescita. Le prime applicazioni (Boccardo ed al., 2007) hanno evidenziato difficoltà di acquisizione, in particolare nell'assicurare un'adeguata copertura stereoscopica. Le immagini presentano ricoprimenti variabili e condizioni di convergenza e deriva piuttosto accentuate, con conseguenti rotazioni e variazioni di scala. Si è quindi ritenuto opportuno testare l'operatore SIFT su alcune coppie di immagini acquisite con la camera RICOH GR (3264x2448 *pixel*) montata sull'UAV *Pelican*, sviluppato presso il Politecnico di Torino.

In Tabella 4 sono riportati i risultati dei test su una coppia di immagini (331-332, ricoprimento: 50%, Figura 5). Per quanto riguarda SIFT sono state effettuate diverse prove in funzione delle soglie di contrasto della ricerca dei massimi radiometrici variando il valore proposto da Lowe ( $th\_key=0,001$ ). In questo caso i valori standard per la selezione dei punti di massimo nello "spazio delle scale" consentono di estrarre solo 40 features. Riducendo la soglia, si ottiene un netto incremento di punti di interesse (478). ForCC estrae un numero inferiore di punti ma con un'ottima percentuale di coppie non affette da errore grossolano. Il calcolo dell'orientamento relativo ( $L=10$ ), ha messo in evidenza una certa instabilità dei parametri di calibrazione della camera, che non possiede un'ottica fissa. Tali problemi hanno provocato l'eliminazione dei punti ai bordi del modello stereoscopico durante la stima LMS.

$L$	FORSTNER+CC					SIFT				
	$th\_key$	$mth$	$mthok$	%	$P_{max}$ ( <i>pixel</i> )	$Th\_key$	$mth$	$mthok$	%	$P_{max}$ ( <i>pixel</i> )
10	<b>0.8</b>	111	105	<b>94</b>	<b>0.29</b>	<b>0.0005</b>	478	256	<b>53.5</b>	<b>0.06</b>
	<b>0.9</b>	29	27	93	0.13	<b>0.001</b>	40	28	70	0.59

**Tabella 4 - Risultati forniti dagli operatori ForCC e SIFT sulle immagini 331-332**



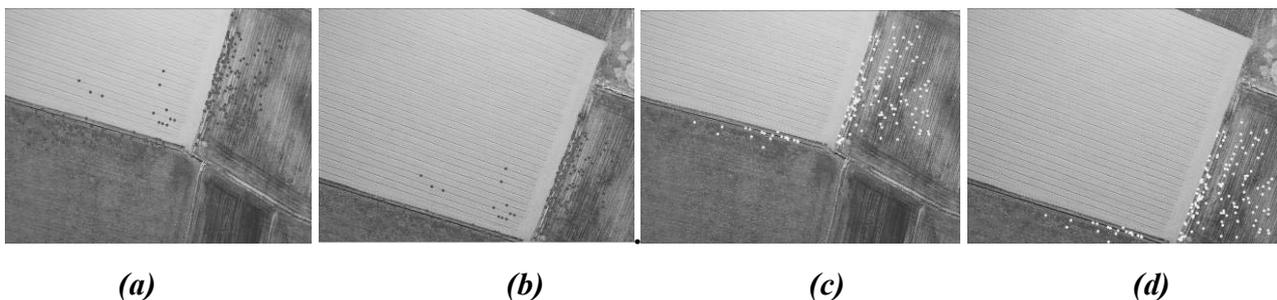
**Figura 5 - Punti omologhi estratti con ForCC (a-b) e SIFT (c-d)**

Un secondo test è stato effettuato sulle immagini 335-336 (Figura 7). In questo caso, la presenza di un campo arato con *pattern* ripetitivi e un lieve effetto di trascinato sull'immagine 335 hanno provocato notevoli difficoltà nel *matching*. Con le soglie standard di SIFT, nessun punto omologo è stato estratto nella zona arata. Per incrementare il numero di coppie si è agito sulla soglia di contrasto sulla funzione  $|D(\hat{x})|$  [0.03, in Lowe 2004] (Tabella 6).

$L$	FORSTNER+CC					SIFT				
		$mth$	$mthok$	%	$p_{max}$ (pixel)	$ D(\hat{x}) $	$mth$	$mthok$	%	$p_{max}$ (pixel)
10	<b>0.8</b>	182	99	54.4	noconv	<b>0.03</b>	205	151	73.6	0.17
	<b>0.9</b>	28	20	71.4	<b>0.18</b>	<b>0.01</b>	745	283	37.6	<b>0.08</b>

**Tabella 6 - Risultati forniti dagli operatori ForCC e SIFT sulle immagini 335-336**

Come si può notare dalla Figura 7, nella zona arata sono state estratte 13 coppie. Il ForCC è invece fallito ( $\epsilon = 0.8$ ), poichè la stima LMS non è stata sufficiente a rimuovere le coppie affette da errore, comportando una mancata convergenza nella stima dell'orientamento relativo.



**Figura 7 - Punti omologhi estratti con SIFT (a-b) e ForCC (c-d)**

### Applicazioni terrestri

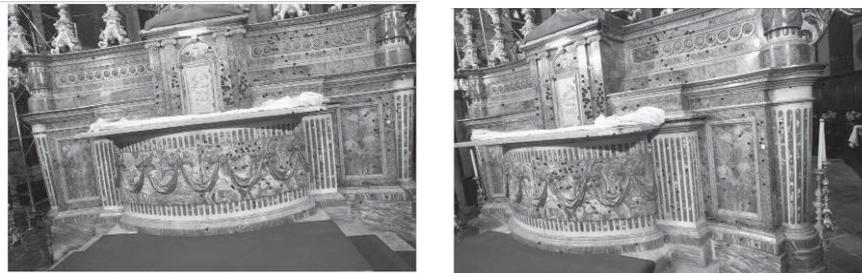
Nelle applicazioni terrestri occorre sovente acquisire immagini secondo schemi di presa non convenzionali al fine di descrivere oggetti complessi. Per questo occorre valutare le potenzialità dell'operatore SIFT nell'orientamento di immagini acquisite con prese convergenti o ruotate fra loro. A tal proposito sono state utilizzate nei test immagini acquisite con una Canon EOS 5D (4368x2912 *pixel*). Durante questi test sono stati utilizzati principalmente i valori delle soglie proposte da [Lowe, 2004], considerando tali valori come ottimali per applicazioni *close range*. Del

resto anche le prove eseguite con valori differenti di soglia non hanno fornito valori significativamente discrepanti. In Tabella 8, sono proposti i risultati ottenuti nell'orientamento di due immagini convergenti (7631-7636) ruotate di oltre 25° fra loro, al variare della soglia di reiezione.

<i>L</i>	FORSTNER+CC					SIFT				
	$\rho$	<i>mth</i>	<i>mthok</i>	%	$P_{max}$ (pixel)	<i>Th_eu</i>	<i>mth</i>	<i>mthok</i>	%	$P_{max}$ (pixel)
5	<b>0.9</b>	256	98	38.2	5.24	<b>2</b>	1500	95	<b>6.4</b>	<b>0.032</b>
8	<b>0.9</b>	256	124	48.4	14.2	<b>2</b>	1500	206	<b>13.8</b>	<b>0.091</b>
10	<b>0.9</b>	256	131	51.1	9.41	<b>2</b>	1500	303	<b>20.2</b>	<b>0.143</b>
15	<b>0.9</b>	256	144	56.2	45.7	<b>2</b>	1500	643	<b>42.9</b>	<b>0.349</b>

**Tabella 8 - Risultati forniti dagli operatori in esame su una coppia di immagini convergenti**

Come si può osservare sulle immagini (Figura 9), le coppie di punti estratte dall'operatore SIFT risultano ben distribuite su tutta l'immagine, evidenziando la bontà della soluzione. Il numero di coppie estratte e i valori di parallasse dimostrano la maggior affidabilità dell'operatore SIFT rispetto al sistema ForCC, troppo sensibile alle deformazioni prospettiche.



**Figura 9 - Punti omologhi estratti con SIFT (L=10)**

Un seconda prova è stata eseguita su una coppia di immagini ruotate in *k* di circa 45° fra loro. Anche in questo caso sono stati considerati diversi valori di soglia di reiezione, mantenendo i valori proposti da Lowe per i *threshold*.

Questa prova dimostra la robustezza dell'operatore SIFT (Tabella 10). L'operatore è in grado di estrarre coppie omologhe anche in presenza di forti rotazioni, al contrario dell'algorithmo ForCC, che non è in grado di fornire una soluzione neppure per valori elevati di *L*.

<i>L</i>	SIFT				
	<i>Th_eu</i>	<i>mth</i>	<i>mthok</i>	%	$P_{max}$ (pixel)
5	<b>2</b>	1353	91	<b>6.7</b>	<b>0.123</b>
8	<b>2</b>	1353	204	<b>15.1</b>	<b>0.158</b>
10	<b>2</b>	1353	344	<b>25.4</b>	<b>0.246</b>
15	<b>2</b>	1353	599	<b>44.3</b>	<b>0.445</b>

**Tabella 10 - Risultati forniti da SIFT su immagini ruotate**

## CONCLUSIONI

Alla luce delle prove eseguite si può affermare che l'operatore SIFT, completamente implementato in ambiente Matlab, è adatto alle operazioni di *feature extraction* anche in condizioni di presa critiche. Si è notato che per le applicazioni aeree (UAV) l'estrazione dei punti di interesse dipende fortemente dai parametri di soglia del contrasto, che devono quindi essere calibrati in funzione dell'oggetto ripreso, delle condizioni di illuminazione e delle casistiche applicative. All'elevata capacità di estrazione di punti di interesse si contrappongono tempi lunghi di elaborazione di ogni immagine e notevoli difficoltà nell'implementazione dell'algoritmo. Durante la fase di stima robusta (LMS) dell'orientamento relativo, la soglia di reiezione gioca un ruolo importante nell'eliminazione di coppie errate. In particolare all'aumentare di  $L$  si riducono le coppie omologhe affette da errore grossolano, con un conseguente decremento dell'accuratezza dell'orientamento.

Allo stato attuale le accuratze fornite da SIFT sono migliori rispetto a ForCC. Si prevede di eseguire un'analisi più dettagliata applicando l'algoritmo di *Least Square Matching* alle coppie estratte mediante *cross-correlation*.

## BIBLIOGRAFIA

Mikolajczyk K., Tuytelars T., Schmid C., Zissermann A., Matas j., Schaffalitzky F., Kadir T., Van Gool L. (2006). "A comparison of affine Region Detectors", *International Journal of Computer Vision* 2006.

Remondino F., (2006). "Image-based modeling for object and human recognition", Doctoral Thesis ETH No.16562, Zurich.

Lowe D. (2004), "Distinctive Image Features from Scale-Invariant Keypoints" *International Journal of Computer Vision* 60(2) pp. 91-110.

Harris C., Stephens M. (1988) "A combined corner and edge detector" *Proc. 4<sup>th</sup> Alvey Vision Conf.*

Rousseau P.J., Leroy A.M. (1987) "Robust regression and outliers detection". *Wiley Series in Probability and Mathematical Statistics*.

Forstner W., Gulch E., (1987) "A fast operator for detection and precise location of distinct points, corners and circular features". *Proc. Inter. Conference on fast Processing of photogrammetric data*.

