

**SUPSI**

# Video processing e computer vision

Fondamenti di Multimedia Processing

Tiziano Leidi

14.12.2017

# Cinematografia

La cinematografia è la scienza e l'arte della fotografia in movimento, in cui la luce visibile viene registrata in maniera ripetuta, chimicamente, per il tramite di un materiale fotosensibile, o elettronicamente, per il tramite di un sensore.



L'impiego di tecniche cinematografiche è frequente in molti ambiti sia commerciali, che scientifici o dell'intrattenimento.

# Videocamere

Nelle videocamere, in maniera analoga alle fotocamere, una lente viene utilizzata per mettere a fuoco ripetutamente la luce riflessa dagli oggetti in immagini sulla superficie fotosensibile all'interno della camera.

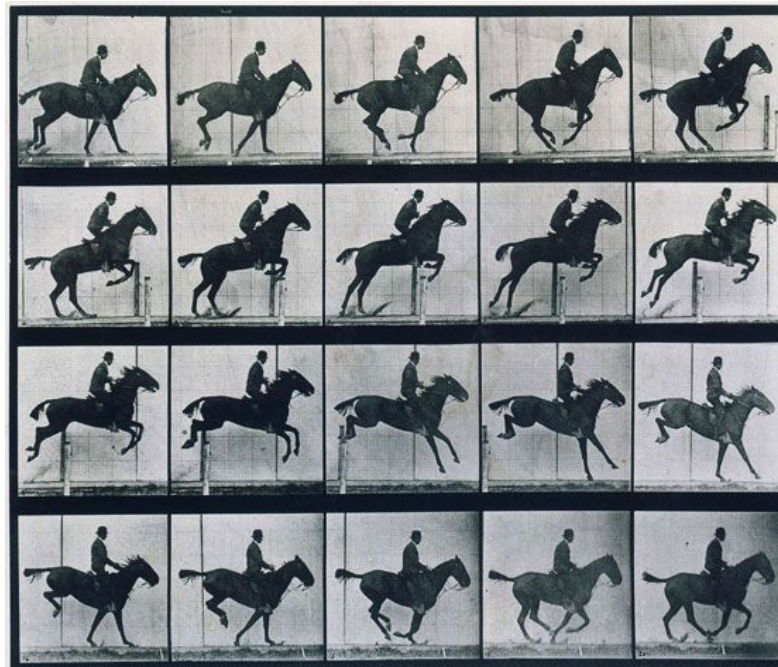
Vengono registrate sequenze di immagini.

Nel caso dell'emulsione fotografica, si tratta di una serie di immagini latenti invisibili sulla pellicola, che viene poi sviluppata chimicamente.

Nel caso d'impiego di sensore elettronico, viene prodotta una scarica elettrica per ogni pixel di ogni immagine, che viene successivamente processata elettronicamente e salvata all'interno di un file video.

# Riproduzione

Le immagini presenti sulla pellicola sviluppata o all'interno del file video vengono successivamente riprodotte in rapida sequenza e proiettate su uno schermo cinematografico o su televisore/monitor, creando l'illusione del movimento.



# Proiettore cinematografico

Un proiettore cinematografico è un dispositivo opto-meccanico per riprodurre film di immagini in movimento proiettandole su uno schermo.

Secondo la teoria del fenomeno beta-phi, il cervello ricostruisce un'esperienza di movimento apparente se gli viene presentata una sequenza di immagini statiche quasi identiche ad un frame rate maggiore di 10/12 immagini al secondo.



# Formati dei film

I proiettori vengono classificati in base alla dimensione del film utilizzato.

I formati più tipici sono:

- per i film amatoriali: 8mm, super 8 e 9.5 mm
- per il cinema: 35mm, 70mm

Oggi esistono inoltre proiettori cinematografici digitali.

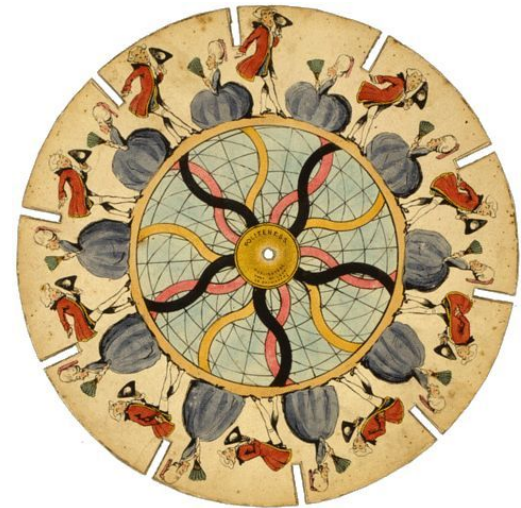
I produttori autorizzati sono: Sony, Barco, Christie and NEC. Sony utilizza la tecnologia proprietaria SXRD, mentre le rimanenti la tecnologia di Digital Light Processing (DLP) sviluppata da Texas Instruments (TI).

## Cenni storici

I primi tentativi risalgono agli anni 1830, in cui sequenze di immagini in movimento venivano create utilizzando dischi rotanti.

L'invenzione della prima camera capace di registrare sequenze di immagini continue è del 1845, utilizzata per la meteorologia.

È a partire dal 1873 che si cominciano a vedere i primi casi di cinematografia vera e propria, applicata negli anni che seguono sia a sperimentazioni scientifiche, sia nell'ambito dell'intrattenimento.



## Cenni storici

A Parigi nel dicembre del 1895, Louis and Auguste Lumière perfezionarono il cinematografo, un'apparecchiatura per registrare, stampare e riprodurre films.

I fratelli Lumière furono i primi a presentare proiezioni di film a pagamento ad un pubblico più grande di una singola persona. Nel 1896, i primi cinema furono aperti in Francia (Parigi, Lione, Bordeaux, Nizza, Marsiglia), Italia (Roma, Milano, Napoli, Genova, Venezia, Bologna, Forlì), Bruxell e Londra.

In seguito, l'evoluzione fu massiccia, con l'introduzione del film a colori e sonoro fra il 1900 e il 1920 e l'avvento del digitale a partire dagli anni '80.



# Funzionamento e storia delle videocamere

[https://www.youtube.com/watch?v=\\_zpZ63\\_U-ws](https://www.youtube.com/watch?v=_zpZ63_U-ws)

# Televisione

La televisione è un dispositivo che combina un tuner, un display e degli altoparlanti per permettere la visione di trasmissioni televisive.

I primi modelli avevano un funzionamento di tipo meccanico, sfruttando un tubo a neon, un disco rotante e un vetro d'ingrandimento. In seguito la tecnologia si sviluppò principalmente su tubi a raggi catodici.



## Cenni storici

A livello sperimentale, le prime televisioni furono sviluppate alla fine degli anni '20, ma raggiunsero il mercato solo dopo diversi anni. È con la fine della seconda guerra mondiale che il broadcasting televisivo in bianco e nero divenne popolare negli Stati Uniti e in Inghilterra.

Negli anni '50, la televisione era lo strumento principale per influenzare l'opinione pubblica.
















Nella metà degli anni '60 fu introdotto il broadcasting a colori.

A partire dagli anni '90 l'evoluzione accelerò con l'introduzione della televisione digitale.

Nel 2013, il 79% delle case del nostro pianeta aveva una televisione.

# Cenni storici

## THE EVOLUTION OF TELEVISION

 <p><b>1928</b></p> <p><b>OCTAGON</b></p> <p>General Electric made the Octagon in 1928 as part of their experimental TV program in Schenectady, New York. The first TV drama, the Queen's Messenger, was produced in September of that year by GE.</p>	 <p><b>1930</b></p> <p><b>BAIRD</b></p> <p>The Baird Television was made by Plessey in England from 1930 through the early 30s. It was the first television receiver sold to the public.</p>	 <p><b>1936</b></p> <p><b>EMYVISOR &amp; COSSOR</b></p> <p>In the year 1936 two TV was invented the first was the Emyvisor which picture in black &amp; white. the next TV was the Cossor which not only show picture in black and white but also in color.</p>	 <p><b>1938</b></p> <p><b>MARCONI</b></p> <p>Here's a 1938 Marconi 707 Television &amp; All Wave Radio Receiver. Measuring 26" x 19" x 19" and weighing more than 100 pounds, the set was actually considered to be somewhat compact in its day, and though its 7 inch screen would be regarded as miniscule by contemporary standards, in 1938 it was not insubstantial.</p>	 <p><b>1939</b></p> <p><b>RCA</b></p> <p>RCA introduced television to the American public at the 1939 World's Fair. Before the fair, they published a brochure for their dealers to explain television.</p>	 <p><b>1946</b></p> <p><b>RCA</b></p> <p>The RCA 630TS television became an immediate hit when it was introduced in 1946, right after World War II.</p>	 <p><b>1948</b></p> <p><b>MOTOROLA</b></p> <p>Motorola's "Golden View" was the most popular 7-inch television in the late 1940s and early 1950s. It came in both tabletop and portable cabinets and it was one of the cheapest TVs available at the time.</p>	 <p><b>1949</b></p> <p><b>RAYTHEON</b></p> <p>The Raytheon M-1101 is an American TV set manufactured in Raytheon's Belmont Radio plant in Chicago on October 1949, the CRT face was more or less masked to give a rectangular appearance), this style of TV is known as "porthole", like the "windows" on a ship.</p>
 <p><b>1953</b></p> <p><b>SHARP</b></p> <p>Sharp started producing as the first Japanese television in mass production. The 14-inch TV was the standard in the first Japanese households for years. With its wooden frame, it precisely met the design aesthetic taste of the fifties.</p>	 <p><b>1958</b></p> <p><b>PHILCO</b></p> <p>Is this the ultimate TV? Love it or hate it, the Philco Predicta television is unarguably one of the design icons of the 20th Century.</p>	 <p><b>1962</b></p> <p><b>MEIDENSHA</b></p> <p>Meidensha TV's were really contemporary in style and design. The wooden frame and high voltage tubes were considered as a great combo back then.</p>	 <p><b>1973</b></p> <p><b>PHILCO-FORD</b></p> <p>1973 Philco-Ford - Model B450ETG - One of the last 'vacuum tube' sets. It was in this time period that the American television set industry migrated to a transistorized TV chassis.</p>	 <p><b>1998</b></p> <p><b>SONY</b></p> <p>The Sony T.V was created in 1998, it was the first television that had a built in VCR and DVD player. The Sony t.v had better picture and a lot more channel with color.</p>	 <p><b>2007</b></p> <p><b>SAMSUNG</b></p> <p>Samsung emerged as one of the largest flat panel TV producer worldwide. Samsung also introduced a ten-millimeter thick only, 40-inch LCD television panel for the first time too.</p>	 <p><b>2014</b></p> <p><b>SAMSUNG</b></p> <p>Samsung started selling commercial curved smart TVs. In the IFA2014, Samsung also displayed the first bendable TV with 5,120 x 2,160 resolutions.</p>	

# Supporti per il salvataggio

L'evoluzione negli anni di vari tipi di supporti per il salvataggio di filmati e degli associati registratori/riproduttori video ha reso possibile l'immagazzinamento di materiale cinematografico e televisivo per la riproduzione a casa.

I formati più conosciuti: Betamax, cassette VHS, DVD e dischi Blu-ray ad alta definizione.

# Tecnologie per display

- CRT (Cathode Ray Tube): è un tubo sottovuoto contenente uno o più emettitori di elettroni combinato con uno schermo fluorescente. Per creare le immagini accelera e deflette i fasci di elettroni sullo schermo.
- DLP (Digital Light Processing): è una tecnologia per video proiettori che sfrutta micro-specchi digitali. Viene normalmente impiegata per i proiettori utilizzati negli ambiti commerciali o dell'insegnamento.
- PDP (Plasma Display Panel): è un pannello piatto che utilizza piccole celle contenenti gas ionizzato caricato elettricamente. Utilizzato principalmente per grossi schermi.

# Tecnologie per display

- LCD (Liquid Crystal Display): è un pannello piatto che utilizza cristalli liquidi modulati dalla luce. I cristalli liquidi non emettono direttamente la luce. Per produrre la luce vengono quindi utilizzati dei riflettori o delle backlight (tipicamente delle luci fluorescenti nelle LCD-TV o dei LED nelle LED-TV).
- OLED (Organic Light-Emitting Diode): è una tecnologia che per emettere la luce sfrutta LED, che come elemento elettro-luminescente hanno uno strato di materiale organico capace di produrre differenti tipi di luce in risposta alla sollecitazione elettrica ricevuta.

# Risoluzione

Nel cinema digitale, la risoluzione è normalmente rappresentata dalla quantità di pixel orizzontali. Ad esempio: 2K ( $2048 \times 1080$  quindi 2.2 megapixels) o 4K ( $4096 \times 2160$  quindi 8.8 megapixels).



# Risoluzione

Nelle televisioni e nei monitor le risoluzioni più frequenti sono:

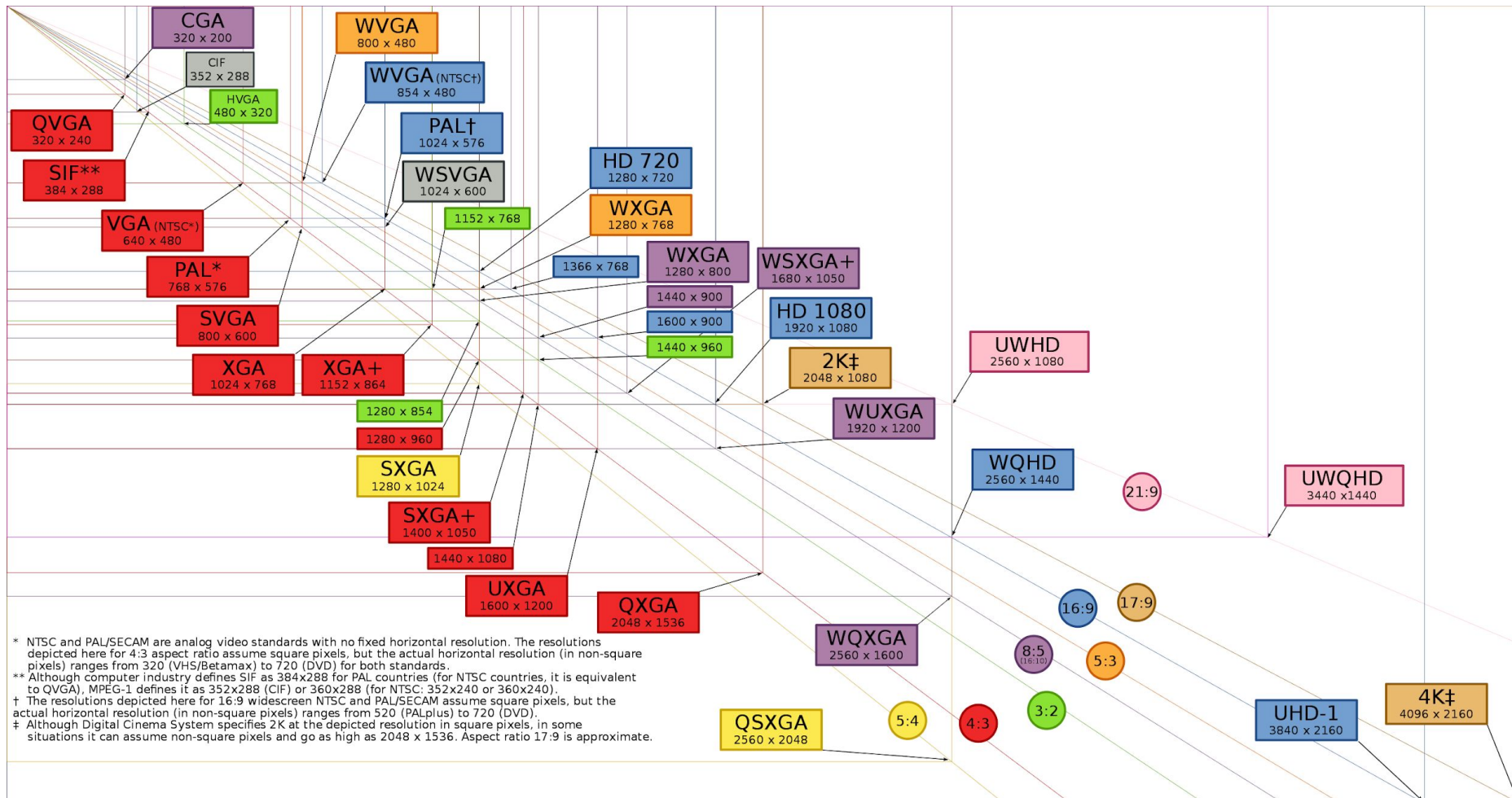
- SD (Standard Definition): 576i, sistema PAL/SECAM Europeo con 576 linee interlacciate; 480i, sistema NTSC Americano con 480 linee interlacciate. Le videocassette VHS possono essere considerate SD, hanno una risoluzione approssimativa di 480i/576i×360.
- HD (High Definition): 720p - 1280×720p progressive scan quindi 921'600 pixels; 1080i - 1920×1080i interlacciato quindi 1'036'800 pixels; 1080p - 1920×1080p quindi 2'073'600 pixels.
- UHD (Ultra High Definition): 4K - 3840×2160 progressive scan; True 4K - 4096×2160; 8K - 7680×4320; True 8K - 8192×4320.

# Aspect ratio

L'aspect ratio di un'immagine descrive la proporzione fra la sua larghezza e la sua altezza.

- 4:3 è il formato utilizzato dall'invenzione della cinematografia, poi nei televisori e nei monitor. I film a 35 mm del cinema muto erano in 4:3.
- 16:9 è il formato standard internazionale dell'HDTV e del DVD.
- 21:9 è un formato cinematografico recente utilizzato oggi negli ultra-wide monitors.

# Risoluzione e aspect ratio



# Frame rate

Il frame rate, espresso in frames al secondo (FPS), è la frequenza con cui le immagini vengono mostrate in un filmato.

La sensibilità temporale della visione umana varia a dipendenza delle caratteristiche dello stimolo visivo ed è differente per ogni individuo.

Di norma, le immagini vengono mostrate a velocità costante. Nel cinema la velocità tradizionalmente impiegata è 24 frames per secondo, nella televisione PAL/NTSC il framerate è di 25/30. Oggi in diversi contesti si usano frame-rate più alti.

# Video processing

In maniera simile all'image processing, nel video processing si utilizzano tecniche di filtraggio dei segnali applicandole a streams di tipo video.

Le tecniche di video processing vengono impiegate in svariati ambiti, compresi quelli della televisione e del cinema e comprendono, fra le altre: modifica dell'aspect ratio; zoom digitale; aggiustamento di brightness, contrast, hue, saturation, sharpness, gamma; conversione del frame rate; aggiustamento del colore; riduzione del rumore; tecniche di upscaling; e molto altro.

# Video processing

Nel video processing vengono impiegate una moltitudine di tecniche dell'immagine processing.

Tutte le tecniche attuate a livello di singola immagine vengono dette tecniche di intra-frame processing.

Invece, le tecniche che sfruttano l'informazione temporale esistente fra più di una immagine in sequenza vengono dette tecniche di inter-frame processing.

# Formati di codifica e compressione video

I formati di codifica e compressione video sono formati di rappresentazione dei dati usati per il salvataggio e nella trasmissione di contenuti video digitali.

Esempi di formati video sono: MPEG-2 Part 2, MPEG-4 Part 2, H.264 (MPEG-4 Part 10), HEVC e RealVideo RV40.

I formati sono definiti in una specifica. Un'implementazione hardware o software di un formato video è un video codec.

Esempi di video codec sono Xvid e OpenH264, che sono due delle varie implementazioni di MPEG-4 Part 2, rispettivamente H.264.

# Formati container per il multimedia

I contenuti video codificati in un particolare formato video vengono normalmente combinati con una stream audio tramite un formato container per il multimedia come AVI, MP4, FLV, RealMedia, Matroska, o QuickTime.

Di conseguenza, l'utente non manipola files di tipo H.264 ma piuttosto .mp4, che è un container MP4 con video codificato in H.264 e audio di norma in formato AAC.



# Compressione video

I filmati video vengono generalmente compressi utilizzando compressioni di tipo lossy, capaci di comprimere maggiormente dei compressori lossless.

Esistono compressori disegnati esplicitamente per compressione lossy rispettivamente lossless, mentre alcuni formati video come il Dirac e l'H.264 permettono entrambe le tipologie.

Una sottoclasse di formati di codifica video relativamente semplice è quella dei formati intra-frame, in cui la compressione viene applicata esclusivamente alle singole immagini senza approfittare della correlazione tra immagini successive. Un esempio è il Motion JPEG.

# MPEG

Il Moving Picture Experts Group (MPEG) è un gruppo di lavoro di autorità che è stato formato da ISO e IEC per stabilire uno standard per la compressione e la trasmissione audio e video. Lo standard MPEG consiste in una moltitudine di parti, ognuna delle quali copre un determinato aspetto della specifica. Negli anni lo standard si è evoluto, di conseguenza alcune parti possono rivelarsi in parziale sovrapposizione con quelle precedenti.

## H.264 - MPEG-4 Part 10

H.264 anche chiamato MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC) è un formato standard orientato a blocchi e basato sulla compensazione del movimento per la compressione video. Dal 2014 è uno dei formati più utilizzati per la registrazione e la compressione di contenuti video. Supporta risoluzioni fino a  $4096 \times 2304$ , incluso il 4K UHD.

# Inter-frame prediction

I filmati video contengono molta informazione ridondante spaziale, ma soprattutto temporale. Negli algoritmi di compressione come H.264 viene sfruttata questa ridondanza con tecniche di inter-frame prediction, in particolare avvantaggiandosi della correlazione tra immagini vicine nella sequenza.

# Inter-frame prediction

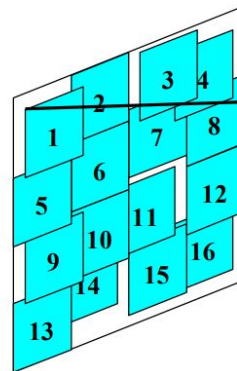


Previously Coded Frame  
(Reference Frame)

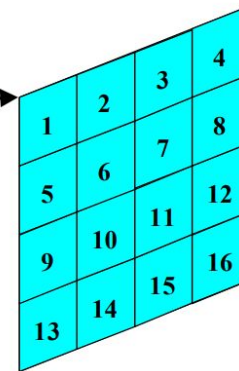


Current Frame  
(To be Predicted)

- Block-matching overview:
- 1) Split current frame into 16x16-pixel blocks
  - 2) Find best match for each block from prior frame



*Reference Frame*



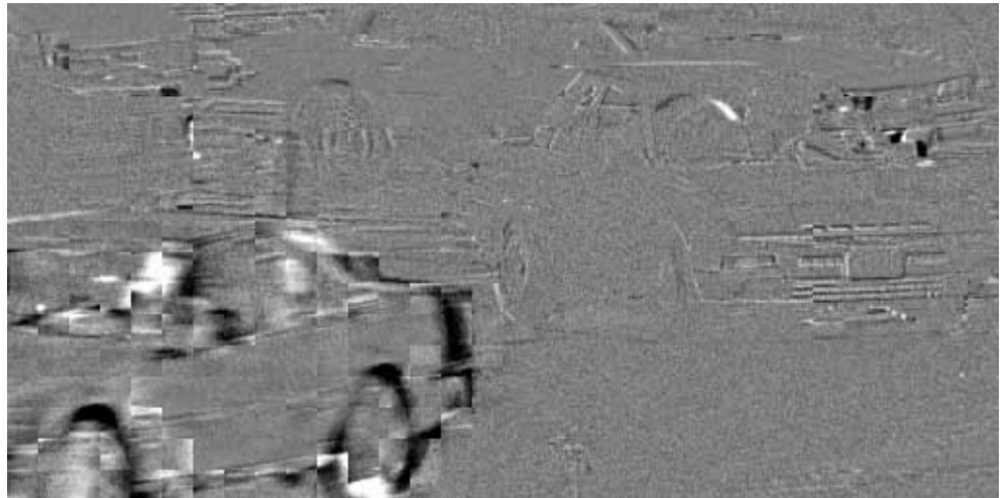
*Predicted Frame*

# Inter-frame prediction

Prediction of  
Current Frame



Prediction Error  
(Residual)



# Motion JPEG

Il Motion JPEG (M-JPEG o MJPEG) è un formato di compressione video in cui ogni frame è compresso separatamente come immagine JPEG.

Viene principalmente utilizzato nelle videocamere, in particolare nelle webcam.

M-JPEG usa uno schema di compressione intraframe. È quindi meno computationally intensive delle soluzioni che applicano interframe prediction.

# Motion JPEG

Le compressioni come MPEG2 e H.264/MPEG-4 AVC, possono raggiungere rapporti di compressione di 1:50 o anche meglio. M-JPEG ha un'efficienza limitata a 1:20 o meno, ma necessita di meno risorse di processing (tempo CPU e memoria).

Inoltre, l'M-JPEG tollera cambiamenti veloci nella videostream, cambiamenti per cui gli schemi di compressione interframe possono presentare perdite di qualità inaccettabili.



# Computer Vision

Le soluzioni di computer vision hanno l'obiettivo di emulare artificialmente le capacità del sistema visivo umano.

La computer vision comprende tecnologie e metodi per acquisire, modificare ed analizzare singole immagini o sequenze di immagini digitali, con lo scopo ultimo di estrapolare informazioni per poter prendere decisioni in maniera automatica (possibilmente senza supervisione umana).

# Computer Vision

I principali campi d'applicazione sono:

- ambito medicale: ha lo scopo di produrre diagnosi medica di un paziente. I dati sono immagini di microscopi, immagini a raggi X, immagini acquisite con tomografia, angiografia, ultrasuoni.
- industria: l'informazione viene estratta con lo scopo di supportare un processo manifatturiero.
- applicazioni militari: individuazione di soldati e veicoli nemici, guida di missili.
- veicoli autonomi: includono sottomarini, robots con ruote, auto, camion, droni e altri veicoli aerei.

# Attività della computer vision

- Riconoscimento: determinare se una classe di oggetti, caratteristiche o attività è presente. Risolto in maniera robusta dall'essere umano, ma non dalla computer vision (caso generale: oggetti arbitrari in situazioni arbitrarie).
- Identificazione: riconoscere un'istanza individuale di un oggetto. Esempi: faccia di una persona, impronta digitale.
- Detection: individuare condizioni specifiche. Esempio: cellule o tessuti abnormi nelle immagini medicali.
- Riconoscimento ottico di caratteri (OCR): identificare caratteri in testo stampato o scritto a mano.
- Image retrieval di contenuti: trovare tutte le immagini con un determinato contenuto comune all'interno di un set.

# Attività della computer vision

- Stima di posizione e orientamento: identificare la posizione o l'orientamento di un oggetto in relazione alla posizione della camera. Esempio d'applicazione: soluzione di computer vision a supporto di un braccio automatizzato.
- Stima del movimento: identificare la direzione e la velocità degli oggetti.
- Tracking: seguire i movimenti degli oggetti.
- Ricostruzione della scena: data una o più immagini di una scena, rispettivamente un video, ricostruire un modello 3D (semplificato o più complesso) della scena.

# Struttura di un sistema di computer vision

Un sistema di computer vision è composto da:

- acquisizione dell'immagine: produzione di immagini digitali tramite sensori o altri tipi di strumenti (ad esempio radar). Le immagini o i filmati possono essere di tipo bidimensionale o 3D.
- Pre-processing: prima dell'estrazione dell'informazione, può essere necessario trattare le immagini per la rimozione del rumore, l'adeguamento del contrasto, l'adattamento della scala di rappresentazione, ecc.
- Estrazione delle feature: estrazione di caratteristiche a vari livelli di complessità. Ad esempio: linee, contorni, spigoli, blob di punti, forme, textures, movimento.

# Struttura di un sistema di computer vision

- Detection/Segmentation: decisione su quali punti o regioni dell'immagine sono rilevanti per i successivi passi di processing.
- Processing di tipo high-level: a questo punto le immagini sono state segmentate in piccoli set di dati, ad esempio set di punti o regioni, supposti di contenere un oggetto o caratteristica. Il processing successivo si occupa di:
  - verificare che i dati soddisfino determinate assunzioni.
  - stima di determinati parametri specifici (posizione, dimensione, ...).
  - classificazione degli oggetti in categorie.

# Background subtraction

La tecnica di background subtraction, anche conosciuta come foreground detection, ha l'obiettivo di estrarre elementi in primo piano (ad esempio esseri umani o automobili) all'interno di un filmato.

Viene utilizzata principalmente nel caso di acquisizione di immagini da videocamere statiche.

Il principio di base è quello di identificare l'oggetto in movimento confrontando il frame corrente con un frame di riferimento, di norma chiamato "background image", o "background model".

# Background subtraction

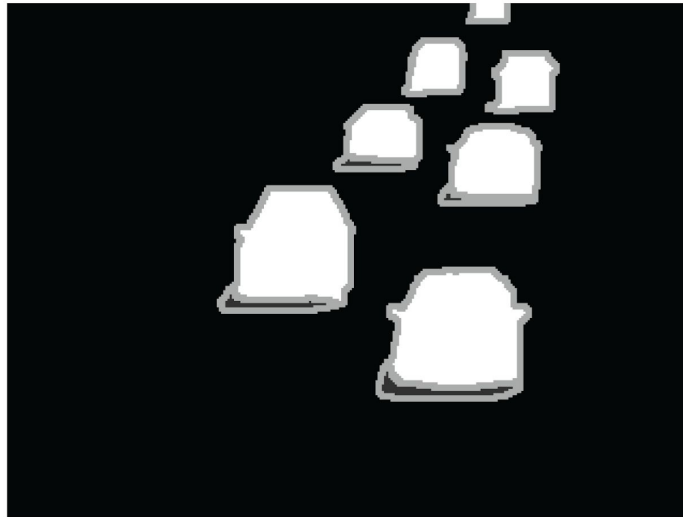
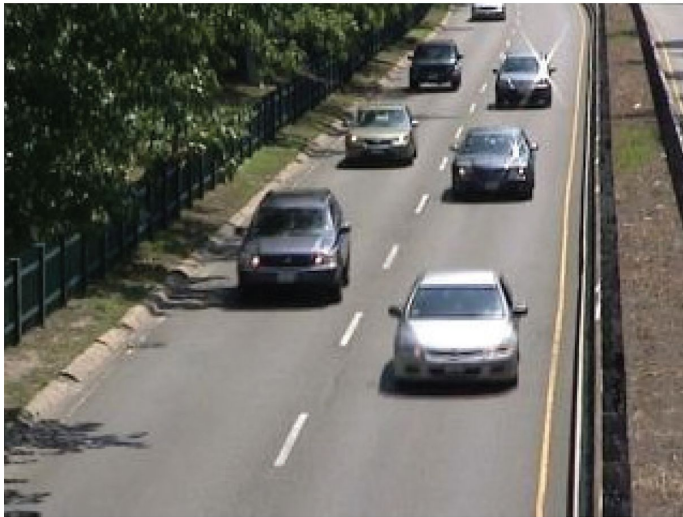
La tecnica di background subtraction si basa su un'ipotesi che spesso non è applicabile ad ambienti reali. Un algoritmo di background subtraction robusto dovrebbe essere capace di gestire cambi di luce, effetti del vento o della pioggia, elementi di ingombro e modifiche della scena a lungo termine.

Esistono varie tecniche di background subtraction:

- Frame differencing
- Mean filter
- Gaussian average



# Background subtraction



# Frame Differencing

Segmenta gli oggetti utilizzando la sottrazione d'immagine per ogni pixel in  $I(t)$  con il corrispondente pixel nella stessa posizione dell'immagine di background  $B$ .

$$P[F(t)] = P[I(t)] - P[B]$$

L'immagine risultante presenterà intensità luminosa solo per i pixels che hanno subito modifiche nei due frames.

Quest'approccio funziona solo se i pixel di background sono statici. Solitamente sulle immagini risultanti viene applicato un threshold.

$$|P[F(t)] - P[F(t + 1)]| > \text{Threshold}$$

# Video motion tracking

Lo scopo delle tecniche di video tracking è di associare gli oggetti in frames video consecutivi. L'associazione può rivelarsi complessa se gli oggetti si muovono rapidamente rispetto al frame rate.

Un'altra situazione che accresce la complessità del problema è nel caso di cambi di direzione. In queste situazioni possono essere impiegati modelli di moto che descrivono e quindi vincolano i cambiamenti di moto accettabili.

# Video motion tracking

Ci sono due componenti principali di un sistema di visual tracking:

- rappresentazione e localizzazione del target,
- filtraggio e associazione dei dati.

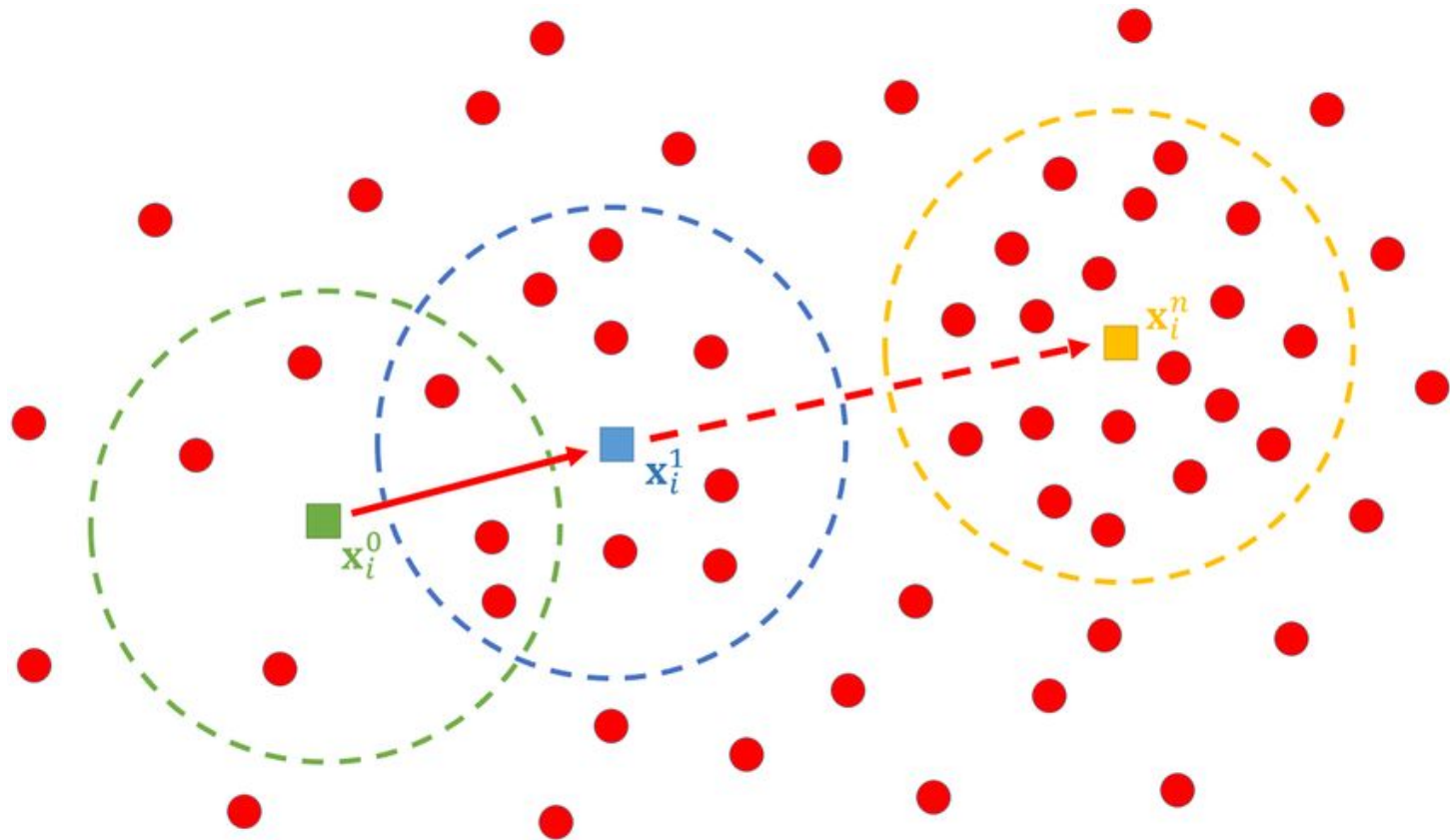
Per la prima parte vengono usate tecniche di kernel-based tracking, come il mean shift, o di contour tracking. Per la seconda parte vengono applicati Kalman filter o Particle filter.

# Mean Shift

Il mean shift è una tecnica di analisi non parametrica per identificare i maxima (detti modes) di una funzione di densità. Viene detto algoritmo di mode-seeking.

Ha un comportamento iterativo che massimizza una misura di similitudine. Viene impiegato in prevalenza per l'analisi dei clusters.

# Mean Shift



# Mean Shift

La variante più semplice impiegata per il visual tracking crea una confidence map della nuova immagine, basandosi sugli istogrammi dei colori dell'oggetto presente nell'immagine precedente. Utilizza l'algoritmo di mean shift per identificare il picco in prossimità della vecchia posizione dell'oggetto.

La confidence map è una funzione di densità di probabilità. Ad ogni pixel della nuova immagine viene assegnata la probabilità che pixels di quel colore siano presenti nell'oggetto.

L'idea viene estesa in varianti come il kernel-based object tracking e il CAMshift.