

Percorso Autonomo Autorizzato

Title (Titolo)	Image and signal processing for multimedia applications and robotics (Elaborazione di immagini e segnali digitali per applicazioni multimediali e robotiche)
Chief (Referente responsabile)	(DEIB, PoliMi): prof. Marco Marcon
Supporting Coordinators (Altri referenti)	(DEIB, PoliMi): prof. Umberto Spagnolini, prof. Stefano Tubaro
Scientific collaborations and partnerships (Collaborazioni scientifiche nazionali ed internazionali)	<ul style="list-style-type: none"> • (DEIB, PoliMi): prof. Fabio Rocca, prof. Claudio Prati • (DMEC, PoliMi): prof. Alfredo Cigada, D.R. Remo Sala • (DAER, PoliMi): prof.ssa Michèle Lavagna • (Dip. Elettronica, PoliTo): prof. Gabriella Olmo, prof. Enrico Magli • (Dip. Elettronica, UniBs): prof. Riccardo Leonardi • (Dip. Elettronica, UniUd): prof. Roberto Rinaldo • (Dip. Elettronica, UniSi): prof. Mauro Barni • (Queen Mary University, London): prof. Andrea Cavallaro • (Imperial College of London): prof. Patrick Naylor, prof. Pier Luigi Dragotti • (Istituto Superior Tecnico, Lisbon): prof. Fernando Pereira • (Humboldt University Berlin): prof. Peter Eisert • (EPFL, Lausanne): prof. Martin Vetterli • (Nuremberg-Erlangen University): prof. Walter Kellermann, prof. Rudolf Rabenstein • (Uni Télécom ParisTech) : Béatrice Pesquet-Popescu • (Dept. of EE-Systems, USC, L.A.): prof. Antonio Ortega • (Mississippi State University): prof. James E. Fowler • (Boston University): prof. Janusz Konrad • (New Jersey Inst. Of Tech., NJ): prof. O.Simeone, A.Haimovich • (U.C. Berkeley): prof. Kannan Ramchandran
Descrizione ed obiettivi	<p>Immagini e video giocano un ruolo sempre più importante sia nella comunicazione interpersonale che nell'interazione uomo-macchina. Si pensi solo al fatto che a fine 2014, circa 2 miliardi di immagini erano caricate ogni giorno solo sui social network mentre su Youtube® erano caricate 300 ore di video ogni minuto. Allo stesso tempo le informazioni visuali hanno un ruolo sempre più importante in un sempre maggiore numero di applicazioni: robotica, controllo produzione industriale, veicoli autonomi, diagnosi medica... L'integrazione tra segnali multimediali provenienti dalle più disparate tipologie di sensori e sistemi avanzati di elaborazione digitale consente di avere una descrizione molto dettagliata dei segnali provenienti dall'ambiente circostante dai quali è possibile estrarre informazioni di alto livello che consentono, ad esempio, di movimentare un robot in contesti non noti a priori. Nella comunità internazionale si assiste ad un fermento straordinario legato a queste tecnologie, con una notevole produzione brevettuale oltre che scientifica.</p> <p>Ciò permette da una parte la nascita di fenomeni completamente nuovi, si pensi ad esempio a YouTub®e (che ha generato una modalità creazione-fruizione dei contenuti video completamente nuova) o al file-sharing basato su reti peer-to-peer (che sta forzando una completa ridefinizione delle modalità per gestire il diritto d'autore), ma dall'altra cambia completamente gli oggetti quotidiani: si pensi a tablet e smart-phones dotati di tele/foto-camera che hanno praticamente sostituito le macchine fotografiche. Nel frattempo stanno nascendo nuovi servizi come la ricerca all'interno di archivi fotografici fatta attraverso ricerche visuali (cerca tutte le immagini connesse ad una di riferimento, cerca tutte le immagini in cui è presente il volto della nonna, ...) e nuove domande di tipo etico: quanto posso credere alla realtà e veridicità di una fotografia o di un video quando essi possono essere manipolati in maniera completamente realistica? Al tempo stesso questa rivoluzione ha creato una serie completamente nuova di applicazioni, ambiti di ricerca/lavoro,</p>

	<p>professionalità. Da una parte la produzione di programmi televisivi e cinematografici combina in modo “seamless” contenuti reali e virtuali, in questo le competenze di analisi ed elaborazione delle immagini sono essenziali. Applicazioni quali Boujou, realizzato da esperti di elaborazione di immagini (http://www.vicon.com/boujou), hanno completamente cambiato la postproduzione cinematografica. Lo stesso può dirsi per i sistemi di “motion capture”, che catturano i movimenti attraverso l’uso combinato di più telecamere e sono alla base delle moderne tecniche di riabilitazione medica e di animazione di attori virtuali. La rappresentazione di dati 3D e filmati è resa sempre più interattiva grazie a moderne tecniche di computer graphics che consentono un’interazione sempre più immersiva tra l’utente e l’ambiente 3D digitale.</p> <p>Anche in ambito industriale il controllo della produzione è sempre più affidato a sistemi di visione in grado di valutare con tecniche sofisticate la qualità e perfezione di ogni singolo manufatto attraverso l’analisi delle immagini acquisite lungo la catena produttiva potendo, ad esempio, ottimizzare e correggere le azioni eseguite dai robot in funzione di ogni singolo pezzo prodotto. La valutazione della qualità della produzione di banconote, piastrelle, tessuti, ma anche il grado di maturazione della frutta, la qualità della carne sono oggi affidati a sistemi di visione ed elaborazione statistica delle immagini.</p> <p>I sistemi di guida automatica (o assistita) di veicoli utilizzano in modo massiccio informazioni ottenute dall’analisi delle immagini da telecamere multiple poste sul veicolo ed anche la retroazione sui sistemi di controllo della traiettoria viene sempre più affidata a tecniche di automazione e robotica. Anche in ambito ludico le applicazioni di acquisizione ed elaborazione delle immagini si stanno sempre allargando, Microsoft con i sistemi kinect ad esempio propone una modalità di interazione con i videogiochi senza l’uso di joystick basato proprio sull’utilizzo di telecamere per catturare i gesti del giocatore; inoltre gli ambienti virtuali ricostruiti nei videogiochi sono sempre più realistici grazie all’integrazione di tecniche avanzate di computer graphics con le moderne GPU disponibili. Immagini da satelliti multipli, o immagini del sottosuolo rappresentano domini applicativi dove i sensori di immagine sono rappresentati da antenne a microonde su un satellite a bassa-quota oppure una sorgente sismica sul terreno. Le tecniche di elaborazione per le immagini si arricchiscono di una conoscenza approfondita della propagazione delle onde per arrivare a focalizzare immagini con estrema precisione, specie se si pensa alle incertezze dell’esperimento (p.e., errore di alcuni millimetri da un satellite).</p> <p>Una simile evoluzione si riscontra nel settore audio, dove l’analisi dei segnali sta progressivamente rivoluzionando le applicazioni disponibili su dispositivi mobili (smartphone). Sono sempre più diffuse le applicazioni che prevedono un’interazione basata sul linguaggio naturale (ad es. SIRI su piattaforma IOS, oppure in “Google Now”), e cominciano a diffondersi applicazioni che inferiscono lo stato emotivo dell’utente dalla voce e dal movimento, oppure che modellano le preferenze musicali dai contenuti stessi del proprio archivio. Sono disponibili nel mercato dispositivi che consentono di localizzare il parlatore e focalizzare l’attenzione sulla sua voce, per meglio riconoscere e interpretare quanto dice. Un ambito particolarmente promettente dell’elaborazione di segnali audio è quello del <i>music information retrieval</i>, che utilizza tecniche di elaborazione statistica dei segnali per ricavare descrittori a vario livello di astrazione (semantica), che consentono di navigare più agevolmente nel mare di contenuti musicali oggi disponibile sulla rete. Sempre in questo ambito sono disponibili tecnologie per la generazione di stream audio personalizzati (ad es. LastFM, StereoMood, Spotify).</p> <p>Va infine osservato come negli ambiti legati alla produzione video e musicale s’incontrino sempre più frequentemente professionalità che coniugano capacità artistiche e competenze legate all’elaborazione dei segnali.</p>
<p>Description and goals</p>	<p>Images and videos are playing an increasingly role both in interpersonal communication and in human-computer interaction. At the end of 2014 about 2 billion images were daily uploaded on social networks while on YouTube® 300 hours of video are uploaded every minute. At the same time visual information have a growing role in many applications: robotics, industrial control, autonomous vehicles, medical diagnosis. Multimedia signals acquired from a plethora of different sensors together with advanced digital signal processing techniques provide us with a very detailed description of the inputs from the</p>

surrounding environment. Pattern recognition techniques together with robust features descriptors are able to provide high-level information that allow, for example, moving a robot in *a priori* unknown contexts.

The scientific and industrial communities are extremely active in the development and deploying of the multimedia acquisition and processing technologies, with a remarkable production of patents and scientific papers. On one side, tablets and smart-phones have completely replaced the traditional photo/video cameras in our daily usage. At the same time, applications and services have redesigned our life: for example YouTube[®] has redefined the video content acquisition and fruition paradigm, while the file-sharing applications based on peer-to-peer networks have forced a complete revision of the copyright issues for audio/video contents. On the other side, everyday tools and devices changed accordingly: e.g..

Furthermore new services become available, allowing visual search inside photographic archives (search all images similar to a reference object, search all images where the grandmother's face is present, ...) and new ethical questions arise: How can I believe the truthfulness of a photograph or a video when they can be easily manipulated in a realistic way? This revolution generated a completely new series of applications, research areas and jobs: for example, now television programs and movies routinely mix real and virtual contents in seamless way. In fact the joint use of computer vision, image processing and computer graphics techniques allows to seamlessly fuse contents in a realistic way. Applications like Boujou[®], made by image processing experts (<http://www.vicon.com/boujou>), have completely changed the cinematographic postproduction.

Another kind of application field like "motion capture", which captures the movements through the combined use of multiple cameras, built the basis of modern medical rehabilitation techniques and animation of virtual actors. The visualization of 3D data and virtual real time interaction is becoming more and more interactive thanks to modern computer graphics techniques increasingly immersive experience between the user and the 3D digital environment.

In Industrial Quality Inspection, vision systems are continuously increasing their relevance assessing, with sophisticated techniques, the quality of each individual product: the analysis of images acquired along the production chain allows, for example, to discard defective pieces and to optimize robot actions. Compliance with the quality requirements of banknotes, tiles, fabrics, but also fruits or meat portions nowadays is entrusted to vision systems and statistical image processing techniques. Automatic (or assisted) driving systems for vehicles massively use information obtained from the processing of images captured by multiple cameras placed on the vehicles.

Also in gaming applications, image processing and computer graphics techniques are expanding their role: the Microsoft Kinect[®] system, for example, offers the players a full interactive interface without any joystick or wired device. It is based on 3D cameras for automatic gesture recognition of every player and accurate avatars and virtual environments are reconstructed thanks to fast computer graphics algorithms implemented in GPUs.

Satellites or subsurface images represent further application domains where microwave antennas on a satellite or seismic sensors collect data arranged in multidimensional arrays. Then image processing techniques coupled to complex wave propagation models are able to create images that describe with millimeter accuracy the earth surface or its interior.

A similar evolution occurs in the audio field, where audio signal processing is deeply changing our interaction with mobile devices (smartphones, tablets). For example, a series of applications is now providing a natural language interaction with these devices (e.g. SIRI on IOS platform, or "Google Now"), and novel applications are coming out inferring the user's mood by voice and movement, or their musical preferences from the content of his/her archives.

Nowadays there are also off-the-shelf devices able to localize the speaker in a 3D environment and to focus on his voice to better recognize and interpret what he is saying. A particularly promising area of audio signal processing is based on music information

	<p>retrieval, where statistical techniques are able to extract descriptors at different levels of abstraction (semantics), assisting the listener to browse the huge amount of music contents available on the net. Some products are already available for generating custom audio streams (e.g. StereoMood, Lastfm, Spotify).</p> <p>As a last remark, it must be underlined that music and video production areas are increasingly involving people with both artistic skills and deep knowledge on signal processing.</p>
(Piano di studi)	Il Piano di Studi è quello del PSPA di “Statistica” con l’obbligo di alcuni esami caratterizzanti le tematiche dell’acquisizione ed elaborazione di segnali ed immagini.
Study Plan	The programme of studies belongs to the Major in “Applied Statistics” with additional courses in image and video processing. The list of courses can be found in a separate document.
Past MSc theses (Alcune Tesi discusse)	<ul style="list-style-type: none"> • L. Bonacina (Ing. Inf.), <i>3D Models Extraction For Personalized Binaural Audio Applications</i>, 2015 • A. Camarda (Ing. Inf.), <i>Scene Classification for Mountain Landscape Recognition from User Generated Images</i>, 2014 • M. Paracchini (Ing. Mtm.), <i>Localizzazione Robusta dei Punti Salienti del Volto Tramite un Approccio a Regolarizzazione Globale</i>, 2014 • L. Gaborini (Ing. Mtm.), <i>Image tampering detection and localization</i>, 2014 • F. Raimondi (Ing. Mtm), <i>Signal processing for passive seismic and application to Argentière glacier</i>, 2014 • P.F. Piazza (Ing. Mtm.), <i>Analisi congiunta di flussi audio-video per l'estrazione di informazioni sulla volumetria delle scene riprese</i>, 2013 • A. Bolognino (Ing.Tlc), <i>Consensus-based inference methods and cooperative wireless networks</i>, 2013 • G. Soatti (Ing.Tlc), <i>Distributed estimation of multi-link channel in dense cooperative networks</i>, 2013 • G. Sandrini (Ing. Mtm.), <i>Acoustic Imaging in the Ray Space: Application to Environment Inference</i>, 2012 • S. Battaglia, (Ing. Inf.), <i>Individuazione della moto-interpolazione in sequenze video</i>, 2012 • A. Panichella (Ing. Inf.), <i>Identificazione del cambio di bitrate attraverso modelli di stima della qualità in modalità no-reference</i>, 2012 • D. Totaro (Ing. Inf.), <i>Example-base definition of high-level descriptors of musical excerpts</i>, 2012 • D. Rodriguez Salgado (Ing. Inf.), <i>Music recommendation system based on audio segmentation and feature evolution</i>, 2012 • L. Chiarandini (Ing. Tlc.), <i>Automatic audio composing system based on music information retrieval</i>, 2012
Available subjects for a MSc thesis (Tesi disponibili)	<ul style="list-style-type: none"> • <i>Compressed sensing and sparse recovery</i> • <i>Statistical signal processing over sparse geometric graphs</i> • <i>Video tampering detection through compression history analysis, estrazione di parametri di compressione dal bitstream per cercare tracce di compressione multipla</i> • <i>Video tampering detection exploiting steganalytic methods, applicazione di metodi di steganalisi per la detection di modifiche video</i> • <i>Video splicing detection and localization using no reference quality metrics, uso di statistico di metriche no-reference nell'analisi video per riconoscere manipolazioni</i> • <i>Who's best? Image/video quality comparison, ordinare sequenze video o immagini in base alla loro qualità (blurring, logo insertion, crop, etc.)</i> • <i>Odometria visuale e ricostruzione di scena 3D da singola telecamera in moto</i> • <i>Stima diretta delle mappe di visibilità acustica basata su analisi di segnali audio acquisiti con schiere di microfoni</i> • <i>Co-registration and calibration of a low-res 3D ToF camera with a Hi-Res camera</i> • <i>Graph analysis for deformable meshes in space</i> • <i>Feature-based on-the-fly genre classification of audio streams</i>

	<ul style="list-style-type: none"> • <i>Non-invasive affective browsing of audio content</i> • <i>Ottimizzazione delle sensitività spettrali in sensori CMOS multi-canale per imaging multi-banda</i> • <i>Sensori d'immagine CMOS filterless in tecnologia backside</i> • <i>Sensori RGB/IR per rivelazione di immagini 3D</i> • <i>Sensori riconfigurabili ad elevato range dinamico per acquisizione di immagini a colori in condizioni di illuminazione fortemente variabili</i> • <i>Simulazione microelettronica di sensori sensibili al colore basati sulla diffusione</i> • <i>Relazione tra sharpening dei filtri e progetto delle microlenti in sensori sensibili al colore</i> • <i>Studio ed ottimizzazione della intersezione tra lo spazio cromatico visibile e quello di un sensore di acquisizione di immagini con filtri accordabili</i>
Tirocini	Su richiesta degli studenti è possibile attivare tirocini presso enti industriali italiani ed esteri.
Internships	On students' demand it is possible to activate internships with Italian or foreign companies.
Sbocchi lavorativi	<p>Il piano di studi proposto offre una forte qualificazione legata agli ambiti connessi alla acquisizione, elaborazione e fruizione di contenuti multimediali. A valle del conseguimento della Laurea Magistrale gli sbocchi lavorativi sono molteplici sia in ambito nazionale sia internazionale. Essi spaziano dalla grande industria specializzata nella produzione di dispositivi elettronici, ad esempio ST Microelectronics che ha una forte produzione di telecamere per dispositivi mobili e di sistemi per l'elaborazione e miglioramento delle immagini, alle imprese impegnate nello sviluppo di sistemi di visione per i processi industriali. Vi sono inoltre opportunità legate a società impegnate nello sviluppo di applicazioni ad hoc per la post-produzione televisiva e cinematografica. Da non dimenticare inoltre gli ambiti connessi al mondo della sicurezza e videosorveglianza ove la necessità di sistemi automatici per l'analisi di immagini è in continua crescita. L'osservazione della terra ed il telerilevamento sono ambiti nei quali le conoscenze di elaborazione delle immagini hanno vasta possibilità di applicazione dalla ricerca di fonti energetiche alternative alla riduzione di emissioni e controllo dell'ambiente. Per ciò che riguarda l'elaborazione dei segnali audio le opportunità di lavoro spaziano dalla post-produzione musicale, al controllo attivo e passivo del rumore, alla pianificazione dell'impronta sonora di oggetti e beni di consumo (si pensi a quanto il giudizio che diamo di un'auto sia influenzato dal rumore dello scarico o della chiusura di una portiera!).</p>
Job opportunities	<p>The proposed study plan offers a specific and highly qualified education in multimedia contents acquisition, processing and broadcasting. After achieving the MSc many working opportunities will be available both in Italy and abroad. In particular, possible employing contexts range from large enterprises for electronic devices, like ST Microelectronics (where a significant part of their business focuses on embedded cameras for mobile devices and on-chip systems for image and video processing, restoration and enhancement) or companies related to development and deployment of <i>ad-hoc</i> systems for post-production in cinematographic and television movies production.</p> <p>Furthermore, video surveillance and security are other applicative contexts where image processing has a continuously growing role and where students following this plan can find a job.</p> <p>Earth monitoring and remote sensing are further domains where image and signal processing cover a relevant role: e.g. from the research in alternative energies to reduction of harmful emissions and environmental protection. Concerning audio signals processing possible working opportunities range from audio post-production, active and passive noise control, music instruments synthesis and audio footprint design for objects and consumer products (e. g. the noise from an exhaust pipe or from a car door closure can greatly influence our evaluation of the car itself!)</p>