

HUMAN SCREEN

**Simone Marullo, Szymon Stefanek, Riccardo Zambon, Alessandro Pozzebon,
Marco Maggini, Marco Gori**

Università degli Studi di Siena

simone.marullo@student.unisi.it, s.stefanek@gmail.com, riccardo.zambon@unisi.it,
alessandro.pozzebon@unisi.it, marco.maggini@unisi.it, marco@diism.unisi.it

Abstract

Human Screen nasce per sperimentare nuove forme di comunicazione digitale collettiva, tramite la ridefinizione concettuale degli schermi degli smartphone come componenti di una periferica di I/O sociale (la sua esistenza è dovuta alla compresenza fisica di una folla di persone) e *distribuita* (ad ogni dispositivo è delegata la rappresentazione di una ristretta porzione del contenuto multimediale). I dispositivi vengono coordinati come tessere di un mosaico, allo scopo di rappresentare un'immagine statica, un video prescelto o un video generato *on-the-fly* (video performance diretta da un'artista). Il sistema realizzato è caratterizzato da robustezza e versatilità, che lo rendono adatto sia a contesti indoor (e.g. teatri) che outdoor (piazze, stadi, concerti, *etc.*). Il modo migliore per aver subito un'idea di Human Screen è di visualizzare un esperimento condotto presso l'Università di Siena con 100 studenti in aula <https://www.youtube.com/watch?v=zff4HFwbUdo>.

1 Finalità

La tecnologia realizzata nell'ambito del progetto Human Screen si presta ad una varietà di applicazioni, tra cui la possibilità di aggiungere aspetti innovativi a occasioni di intrattenimento tradizionali, sia in eventi di piccole-medie dimensioni (e.g. feste private, locali pubblici) che di grandi dimensioni (e.g. performance, concerti, raduni, coreografie negli stadi).

Nuove situazioni potrebbero invece generarsi proprio a motivo dell'azione di Human Screen, poichè può funzionare simultaneamente come collettore degli input dei partecipanti e come display distribuito dell'output del sistema. Si pensi, ad esempio, ad esperimenti di *crowd wisdom*, la teoria sociologica secondo la quale una massa di individui inesperti in alcune condizioni è in grado di fornire una risposta adeguata a una domanda, più di quanto non siano in grado di farlo esperti selezionati. È possibile anche pensare ad esperimenti di creatività collettiva mediata dall'AI: un'intelligenza artificiale potrebbe raccogliere i pensieri e le sensazioni dei partecipanti e utilizzarli come stimoli per produrre un oggetto artistico da riprodurre su Human Screen.

In generale, Human Screen può essere utilizzato come piattaforma per trasmettere messaggi collettivi, spontanei come nel caso dei *flash mob* oppure finalizzati come nel caso dell'*advertising*.

2 Architettura del sistema

Data la natura del task, il sistema è stato realizzato con un'architettura client-server, dove i client sono gli smartphone e il server è il loro coordinatore. È necessario in ogni caso determinare la localizzazione di ciascun dispositivo nell'ambiente. Sono stati studiati due scenari: localizzazione *lato client* e localizzazione *server-assisted*.

Nel primo caso la localizzazione del dispositivo viene determinata sulla base di informazioni raccolte dai moduli GPS, Bluetooth o WiFi, comunemente presenti sugli smartphone: il server riceve la localizzazione ricavata localmente dal dispositivo.

Nel secondo scenario, che è risultato il più versatile ed efficace ed è stato dunque realizzato nell'ambito del progetto Human Screen, viene implementato un feedback loop con una telecamera. La telecamera è direttamente collegata al server, oppure collegata ad un master client, i.e. un computer tramite il quale la telecamera viene gestita e calibrata.

In generale, i compiti del server consistono nel ricevere connessioni TCP dai client, implementare un protocollo di sincronizzazione, gestire gli *show*, mappare la localizzazione dei client ad uno specifico pixel del media da riprodurre, inviare il flusso del pixel rilevante a ogni client. Nel caso di localizzazione *server-assisted*, il server deve anche elaborare il flusso video dalla telecamera e gestire il processo di *discovery* delle localizzazioni.

Il client invece deve contattare il server, gestire la localizzazione (in proprio o in collaborazione con il server), sincronizzare il proprio orologio, ricevere il flusso del pixel e controllare di conseguenza luminosità e colore del proprio display.

Il modello software implementato si articola in una applicazione server scritta in C/C++ (ottimizzata in modo da poter gestire molti partecipanti), una applicazione client per Android in Java, una applicazione client per iOS in Objective-C; le applicazioni condividono una libreria C++ per implementare le parti in comune.

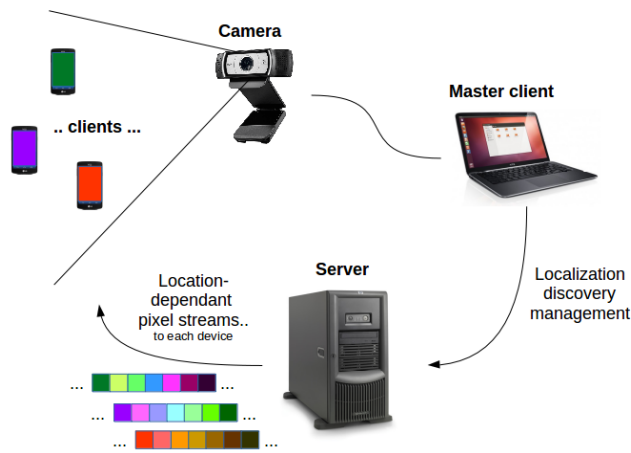


Figura 1: Architettura client-server di Human Screen

3 Altri metodi di localizzazione

La localizzazione tramite GPS è praticabile esclusivamente in contesti outdoor, a motivo della generalmente scarsa ricezione del segnale all'interno degli edifici. Inoltre, i moduli GPS solitamente presenti negli smartphone consentono una localizzazione a meno di un errore tipico di 3-5 metri (da test effettuati sul campo, in uno spazio aperto). Di conseguenza, questo metodo di localizzazione è stato considerato inadatto allo scopo.

La localizzazione tramite Bluetooth, in particolare Bluetooth Low Energy (BLE), è implementabile per mezzo di una classica trilaterazione, sfruttando la potenza del segnale ricevuto - *received signal strength indication* (RSSI) - da un certo numero di *beacon* adeguatamente posizionati. Tuttavia, tale approccio soffre di interferenze e crescente latenza quando la densità di dispositivi in una superficie ridotta diventa notevole [Sugino *et al.*, 2015]; inoltre, il corpo assorbe il segnale quindi potrebbe essere necessario richiedere ai partecipanti di sollevare il dispositivo sopra la testa.

La localizzazione tramite WiFi, implementabile analogamente con trilaterazione (impiegando la misure di RSSI dagli Access Point), permette di ottenere un'accuratezza tipica di 2-2.5m [OnkarPathak *et al.*, 2014]. In aggiunta, operando nella stessa banda del BLE, soffre di problemi analoghi per quanto riguarda scattering, shadowing e interferenze. In ogni caso, la densità di Access Point richiesta per ottenere la localizzazione è superiore alla densità normalmente utilizzata per fornire la sola connettività ai dispositivi.

4 Localizzazione tramite feedback loop visuale e funzionamento di Human Screen

La localizzazione tramite feedback loop visuale, a differenza di BLE e WiFi, non richiede la presenza in loco di una infrastruttura specifica, fatta eccezione per una telecamera (plausibilmente già presente), assicura un alto livello di affidabilità e precisione, minimizza il livello di *behavioral compliance* richiesta ai partecipanti e permette una mappatura immediata del singolo display nella folla con il pixel all'interno del frame da riprodurre.

Il dispositivo contatta il server e si registra come partecipante in un certo *show*; ad ogni dispositivo viene assegnato un identificativo univoco. Viene effettuata la sincronizzazione dell'orologio. A questo punto il dispositivo trasmette visivamente un *discovery pattern*: mostra cioè sul proprio display una sequenza di colori che codificano, secondo un insieme di regole, per il proprio identificativo. Il server, elaborando il flusso video, rileva in una sua specifica posizione una sequenza che rispetta le regole, estrae l'identificativo e memorizza la posizione del dispositivo caratterizzato da quel codice. Quindi trasmette allo smartphone una *timeline*, ossia una sequenza di colori associati ai rispettivi *timestamp* (marcature temporali). Il dispositivo attende il prossimo timestamp e colora il proprio display di conseguenza non appena tale istante viene raggiunto. La riproduzione prosegue fino al termine dello *show*.

È stata verificata inoltre la possibilità di aggiungere un livello audio al contenuto riprodotto, utilizzando gli smartphone anche come speaker distribuiti: alcune caratteristiche morfologiche dell'immagine proiettata vengono utilizzate localmente per modulare in tempo reale sui vari dispositivi volume e frequenza di riproduzione di note o brani musicali, contribuendo ad indurre l'illusione scenica di immagini che fisicamente si muovono nella folla.

Tale processo, coordinato dal server simultaneamente per ogni dispositivo, permette di realizzare la riproduzione distribuita del contenuto multimediale.

5 Flash2

Nel solco di Human Screen è nato il progetto Flash2 (da "flash to (somebody)"), che si è proposto di verificare sperimentalmente se la dotazione tecnologica comunemente presente negli smartphone di ultima generazione fosse sufficiente per supportare un sistema di *instant messaging* locale basato sulla trasmissione ottica di forme e colori tramite i display.

Analogamente a quanto succede nel contesto di Human Screen, un dispositivo trasmette un messaggio, alterando le proprietà visibili del suo display (disegna su di esso pattern colorati). Un altro dispositivo, in questo caso un altro smartphone, elabora il flusso video dalla fotocamera per ricevere il messaggio. È quindi possibile realizzare un sistema di trasmissione con caratteristiche d'uso peculiari (ad esempio la possibilità di *broadcast* ai dispositivi in *line-of-sight* e quindi la sua socialità locale), rivolto a un qualsiasi terminale dotato di webcam (un altro smartphone, un tablet, un computer tradizionale) e utile ad esempio per scambiare dati di contatto, identificativi, richieste e messaggi testuali.

Human Screen e Flash2 possono rappresentare due componenti di un ecosistema di strumenti interattivi, accomunati dall'idea di utilizzare il canale visuale mediato dai display degli smartphone e distinti dal fatto di essere il primo rivolto a una comunicazione collettiva in cui gli utenti sono il supporto di un messaggio condiviso, il secondo rivolto a un ambito più ristretto in cui gli utenti sono gli attori della comunicazione.

6 Circoli creativi

Human Screen può semplicemente considerarsi una tecnologia per la visualizzazione su "schermi umani", ma può contri-

buire all'attivazione di nuovi circoli creativi particolarmente interessanti ed innovativi dal punto di vista artistico. Mentre la soluzione descritta fin qui assume che il video da proiettare sia predefinito e quindi selezionabile nell'applicazione, si può concepire uno schema ben più raffinato ispirato a quanto illustrato in Fig. 2. Nell'architettura illustrata, i patterns vi-

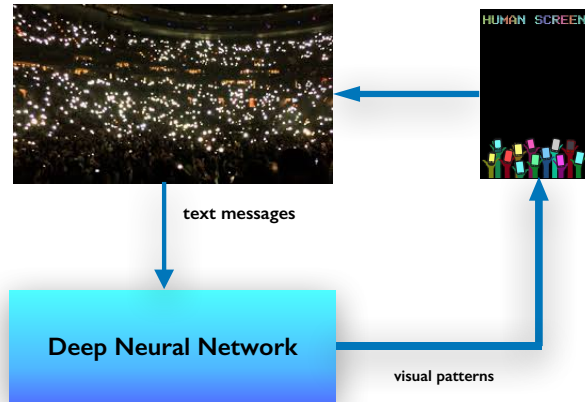


Figura 2: Circolo creativo dell'artista con il modello generativo di video

suali vengono prodotti da un sistema basato su tecnologie di deep learning. Tale sistema acquisisce in ingresso messaggi di testo provenienti dal pubblico e li converte, attraverso un meccanismo generativo, in video. Schemi di simili meccanismi generativi sono discussi in [Marra *et al.*, 2018]. La visualizzazione mediante Human Screen sul pubblico assume in questo schema intriganti aspetti creativi al confine che coinvolgono chiaramente tre "agenti":

- L'artista che, per primo, costruisce il suo messaggio creativo che stimola emozioni nel pubblico;
- Il pubblico che, ricevuto il messaggio scatena le sue emozioni attraverso tweets;
- L'agente autonomo basato su tecnologie di machine learning, che impara a generare video stimolanti sulla base delle informazioni testuali ricevute dal pubblico.

E' chiaro che la convivenza e l'elaborazione creativa in un simile contesto è potenzialmente molto ricca e foriera di schemi elaborati e del tutto inesplorati. Lo schema potrebbe essere ancora arricchito offrendo all'artista la possibilità anche di "disegnare sul pubblico", in modo tale che si possano combinare i suoi pattern visuali con quelli generati dalla macchina, che interpretano il pubblico. Il problema della generazione automatica di pattern visuali può anche seguire schemi di reinforcement learning sotto il controllo dell'artista e, chiaramente, può produrre patterns di alto valore suggestivo grazie all'apposita creazione di un opportuno repository di video e immagini a partire dai quali si procede all'elaborazione. Come già accennato, lo schema è tipico di reti neurali GAN e si può particolarmente arricchire grazie all'opportuno sfruttamento di informazioni vincolari. In altri termini, si può con-

cepire un contesto generativo sotto il controllo di vincoli, che produce rispettando consistenze a partire da grosse collezioni di dati. Le idee di base su simili schemi di apprendimento con consistenza si possono trovare nel cap. 6 di [Gori, 2017].

L'ipotesi di circolo creativo qui discussa è chiaramente difficile da realizzare non tanto per problemi tecnologici, ma per il complesso compito di stimolare artisti in concerti a condurre simili esperienze. Infatti, mentre gli scogli tecnologici appaiono realmente affrontabili, è evidente che lo sviluppo del sistema richiede fasi progressive e che occorre un'organizzazione molto accurata degli eventi per poter ospitare le soluzioni di Human Screen. Una linea di ricerca che appare degna di esplorazione è quella di replicare Human Screen, con il suo corrispondente modello creativo riassunto in Fig. 2, attraverso un modello di social network. L'idea è quella che i fan siano organizzati virtualmente su spalti con coordinate in modo che il meccanismo che conduce alla generazione automatica dei video si possa sperimentare dai testi prodotti anche nell'ambito di una rete sociale. Questa soluzione permetterebbe di sperimentare in modo massivo il sistema di autonomo di generazione ma, soprattutto, potrebbe risultare più semplice per attivare il "circolo creativo" che potrebbe ripetersi poi, come epifenomeno, mediante la tecnologia di Human Screen.

Riferimenti bibliografici

- [Gori, 2017] Marco Gori. *Machine Learning: A Constraint-Based Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2017.
- [Marra *et al.*, 2018] Giuseppe Marra, Francesco Giannini, Michelangelo Diligenti, e Marco Gori. Constraint-based visual generation. *CoRR*, abs/1807.09202, 2018.
- [OnkarPathak *et al.*, 2014] OnkarPathak, Pratik Palaskar, Rajesh Palkar, e Mayur Tawari. Wi-Fi indoor positioning system based on RSSI measurements from Wi-Fi access points: A trilateration approach. *International Journal of Scientific & Engineering Research*, April 2014.
- [Sugino *et al.*, 2015] Kyohei Sugino, Yusuke Niwa, Shun Shiramatsu, Tadachika Ozono, e Toramatsu Shintani. Developing a human motion detector using Bluetooth beacons and its applications. *Information Engineering Express*, 1(4):95–105, 2015.