SENSOR DATA ANALYSIS FOR ADVANCED USER INTERFACES

Filip Chmiel

Master Degree Programme (2), FIT BUT

E-mail: xchmie02@stud.fit.vutbr.cz

Supervised by: Vítězslav Beran

E-mail: beranv@fit.vutbr.cz

Abstract: The paper deals with the designing of interface based on multiple input signals, i.e. multimodal interface. The work also includes an overview of the level at which you can perform data fusion, and different approaches to the layout of the system architecture for multimodal data processing. Final part is the actual design of the system where for the resulting interface was chosen distributed architecture using software agents for processing inputs. As a method for data integration was picked dialogue driven fusion. The result should be an interface for controlling media center and interaction with other devices around the user.

Keywords: Multimodal interface, HCI, dialog driven fusion, Kinect, keyword recognition

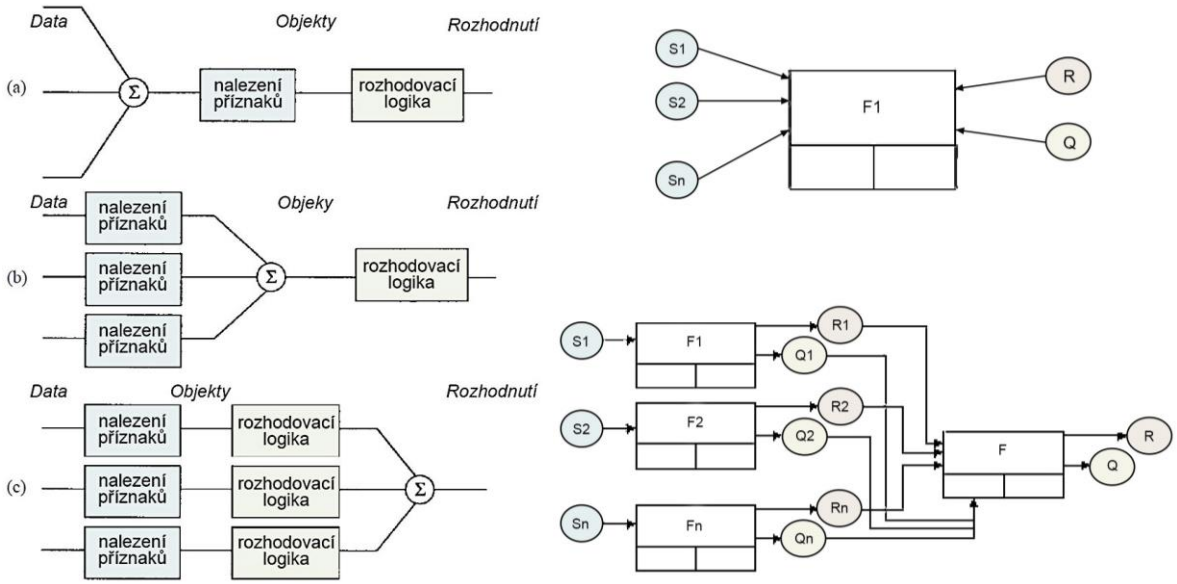
1. ÚVOD

Cílem tohoto příspěvku je přiblížit návrh a realizaci rozhraní založeného na souběžném zpracování více signálů různého charakteru (viz Obrázek 1). Analýza audio či video signálu za účelem získání vstupní akce systému nemusí pro vytvoření rozhraní postačovat. Zvláště, pokud chceme vytvořit uživatelsky přívětivé rozhraní. Proto je práce zaměřená na způsob, jakým vstupní data zkombinovat - tedy použití některé z technik fúze dat.

Obrázek 1: Ukázka možných vstupů pro rozhraní využívající informace z různých zdrojů.

2. FÚZE MULTIMODÁLNÍCH DAT

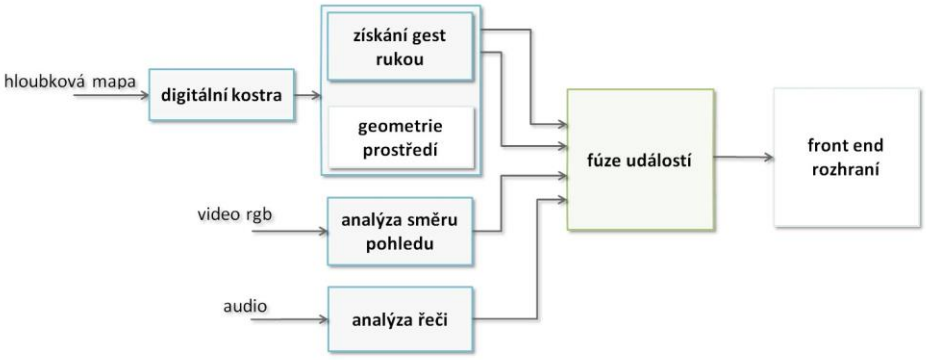
Přístupy k integraci vstupních signálů lze rozdělit jednak podle úrovně, na které k ní dochází, ale také podle rozvržení architektury systému (více viz Obrázek 2). Jelikož navrhovaný systém má být založený na událostech popisujících akce provedené uživatelem, byla vybrána hybridní metoda fúze. Ta kombinuje unifikační a dialogový přístup se zachováním co nejmenší složitosti. Používá gramatiku (fúze unifikací), která pracuje s multimodálními informacemi pomocí definovaných časových a modálních podmínek. Také obsahuje rozhodovací logiku pro filtrování vstupních událostí a rozhodování, zda má být výsledná akce přijata (dialogem řízená fúze) [3].



Obrázek 2: Vlevo: tři úrovně fúze senzorických dat (a) data fusion, (b) feature fusion, (c) decision fusion[2], vpravo: centralizovaná architektura a decentralizovaná s vstupy S, rozhodnutími R a kvalitou Q [1].

3. NÁVRH SYSTÉMU

Na základě informací získaných z dostupných publikací jsem se rozhodl pro hybridní architekturu založenou na samostatném zpracování signálů a jejich následném centralizovaném vyhodnocení (pro bližší představu viz Obrázek 3). Přístup je možno také nazvat jako decision fusion se samočinnými agenty pro zpracování signálu ze senzorů. Jednotlivé vstupy jsou tedy zpracovány specifickými agenty odděleně. Získané informace o akcích uživatele jsou následně předány v podobě události centrálnímu agentovi obstarávajícímu fúzi dat. Ten z příchozích událostí sestavuje komplexnější činnost, kterou má systém vykonat. Výsledek je postoupen grafickému rozhraní pro vykonání či vizualizaci.

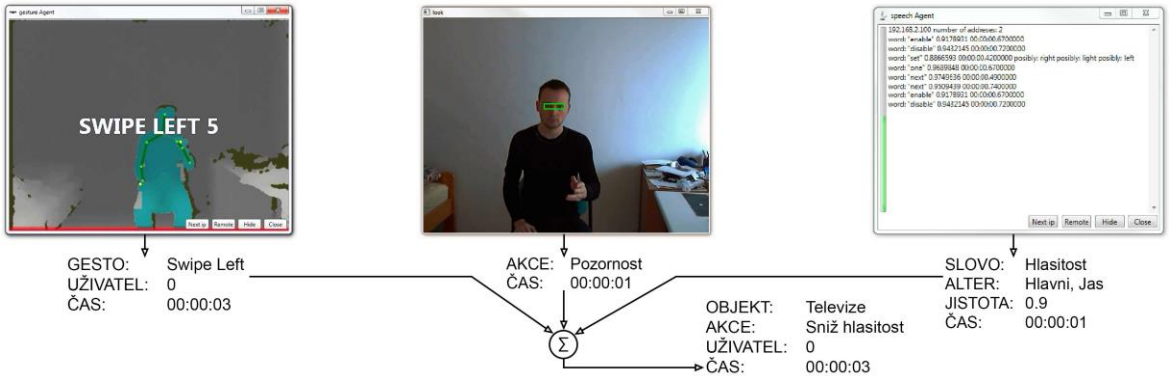


Obrázek 3: Ilustrace návrhu systému pro fúzi dat z více různých zdrojů.

Jednotlivé bloky systému spolu komunikují pomocí protokolu UDP. Specifikovat je lze jako:

a) **Gesta rukou** - díky senzorům (jako Microsoft Kinect) schopným snímat hloubkovou mapu prostoru (navíc přímo podporujícím tvorbu digitální kostry) se v reálném čase sleduje pohyb rukou a následně ze získané trajektorie je určeno, jaké gesto uživatel vykonal. Díky trojrozměrné pozici ruky systém určuje, na který z definovaných objektů uživatel ukázal.
b) **Směr pohledu** – v případě, že uživatel chce komunikovat se systémem (je s ním v očním kontaktu) je směr pohledu aproximován pomocí detekce očí v obraze.
c) **Analýza řeči** - zpracování vstupu v podobě audia je realizováno jako detekce klíčových slov, ta se posílají, spolu s jistotou detekce, centrálnímu agentovi. Tento přístup umožnuje zadávat jednoslovné či kratší slovní příkazy.

d) **Fúze událostí** – pomocí hybridní fúze je z příchozích zpráv sestavena výsledná operace, tedy příkazy pro multimediální centrum (pohyb v nabídce, ovládání přehrávání nebo hlasitosti) a interakci s prostředím (zapínání či vypínání přístrojů).



Obrázek 4: Ukázka implementovaného řešení detekce gest spolu s detekcí směru pohledu a klíčivých slov, následně fúze získaných událostí pro získání výsledné akce

Jak ukazuje Obrázek 4, systém kombinuje gesta a řeč do jednoho příkazu. Modality však nevystupují pouze ve vzájemně doplňující se roli, ale i v konkurenční, kdy příkazy lze zadávat více způsoby za pomocí různých modalit.

Problémem, se kterým se systém musí vypořádat, jsou náhodné pohyby rukou uživatele, které mohou zapříčinit chybné vyhodnocení gesta. Příkazy se vykonávají pouze v případě, je-li člověk v očním kontaktu se systémem (princip podobný mezilidské komunikaci), aby se redukoval počet nechtěných příkazů. Kromě toho je detekce gest prováděna až pro určitou vzdálenost ruky od těla. Takto lze nejen odfiltrovat část nechtěných gest, ale i lépe využít dostupné informace o trojrozměrné pozici uživatele. Nevýhodou zůstává specifické zaměření navrhovaného systému a tedy nutnost dodatečného přizpůsobení pro ovládání jiných aplikací.

4. ZÁVĚR

Pro současné počítače není výpočetní náročnost souběžného zpracování signálu v podobě audia, videa nebo hloubkové mapy problémem. To nabízí další možnosti tvorby uživatelských rozhraní, zejména těch multimodálních. Zde navrhovaný systém může posloužit pro ovládání multimediálního centra s možnosti ovládat i další přístroje kolem uživatele. Díky rozdělení systému do samostatných modulů lze potřebné aplikace umístit na více zařízeních s menším výkonem či změnit použité metody bez nutnosti úprav celého systému. Je ale nutné také počítat s chybovostí senzorů nasazených v běžných podmínkách, a tedy je potřeba dalšího testování na uživatelích.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory grantu FIT-S-11-2 a výzkumného záměru MSM 0021630528.

REFERENCE

[1] SHARMA, Rajeev, Vladimir I. PAVLOVIC a Thomas S. HUANG. Toward multimodal human-computer interface. *IEEE, VOL. 86*. 1998, č. 5.

[2] LALANNE, Denis, Laurence NIGAY, Philippe PALANQUE, Peter ROBINSON, Jean VANDERDONCKT a Jean-François LADRY. Fusion engines for multimodal input: a survey. In: *ICMI-MLMI '09*. New York: ACM, 2009, s. 153-160.

[3] PORTILLO, Pilar Manchón, Guillermo Pérez GARCÍA a Gabriel Amores CARREDANO. Multimodal fusion: a new hybrid strategy for dialogue systems. In: *ICMI '06*. New York: ACM, 2006, s. 357-363.