

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta informačních technologií

Souhrnná zpráva ke smluvnímu výzkumu -  
TESCAN

Brno, 2018

Pavel Najman

# Obsah

Úvod.....	2
RGBD kamery .....	3
Technické řešení .....	4
Vyhodnocení .....	7
Závěr .....	9

# Úvod

S nárůstem výpočetního výkonu se otevírají nové možnosti v oblasti zpracování obrazu a videa v reálném čase. Jednou z otevřených možností je zpracování obrazu s cílem vylepšit interakci člověka s počítačem. Tato nová forma uživatelského rozhraní je založena na sledování uživatele a interpretaci jeho pohybu, gest a chování. Uživatelská rozhraní založená na takovémto sledování se často označují jako přirozená uživatelská rozhraní (angl. Natural User Interface - NUI). Jejich cílem je poskytnout uživateli co nejpřirozenější a nejtransparentnější rozhraní pro komunikaci se strojem. V rámci smluvního výzkumu s firmou TESCAN Brno, s.r.o. (dále TESCAN) bylo navrženo, implementováno a vyhodnoceno uživatelské rozhraní využívající RGBD kameru pro sledování uživatele při interakci se SEM mikroskopem. Technické podrobnosti tohoto zadání jsou uvedeny níže.

## Zadání

Předmětem smlouvy je studie proveditelnosti doplňkového ovládní mikroskopu SEM pomocí RGBD kamery. Toto zařízení bude pevně umístěno na displeji počítače mikroskopu a bude opticky zaměřeno na operátora. Pohyb operátora hlavou směrem k displeji bude ovlivňovat chování softwaru pro řízení mikroskopu.

TESCAN Brno, s.r.o. (dále TESCAN) připraví a do dvou týdnů od podpisu smlouvy předá zhotoviteli potřebné softwarové API v programu TESCAN pro řízení mikroskopu. API bude koncipováno nad protokolem TCP/IP tak, aby bylo možné spojení i vzdáleně po počítačové síti. Fakulta informačních technologií (dále FIT) zajistí sběr dat ze zařízení, filtraci dat, vyhodnocení základních gest uživatele, a předání zpracovaných informací pomocí softwarového API do softwaru TESCAN.

V rámci této studie proveditelnosti budou zpracovány následující techniky ovládní mikroskopu.

1. Řízení dvojrozměrné funkce mírným pohybem hlavy. Např. v případě centrování mikroskopu je zřejmě možné pohybem hlavou vlevo / vpravo / vpřed / vzad měnit hodnotu parametrů, např. centrování. Software FIT bude zasílat informaci o poloze hlavy vlevo/vpravo a vpřed/vzad.
2. Automatické přepínání kontextu okna podle toho, kam se uživatel dívá. Software FIT bude pravidelně předávat informaci o tom, kterou část obrazovky uživatel pozoruje.
3. Navigace v rozlehlém 2D obraze – pohyb hlavy vpřed/vzad bude mapován na funkci zoom v obraze, pootočení hlavy vlevo/vpravo či nahoru/dolů bude sloužit k pohybu daným směrem v panoramatu.

V případě, že se během vývoj objeví nové nápady či metody odlišné od předchozích technik, lze po dohodě smluvních stran zkoumané techniky změnit. Výstupem FIT je praktická demonstrace těchto tří uvedených metod.

Cílem této zprávy je popsat návrh, implementaci a vyhodnocení doplňkového ovládní mikroskopu SEM pomocí RGBD kamery realizovaného v rámci smluvního výzkumu s firmou TESCAN na FIT VUT. Tato zpráva je členěna do 5 krátkých kapitol, které na sebe volně navazují. První kapitola uvádí problematiku navrhovaného rozhraní a popisuje domluvené zadání smluvního výzkumu. Druhá kapitola se věnuje dostupným RGBD kamerám a jejím alternativám použitelných pro realizaci uvažovaného rozhraní. Další kapitola popisuje samotné technické řešení pro jednotlivé body zadání. Čtvrtá kapitola rozebírá provedené uživatelské testy a jejich vyhodnocení. Poslední kapitola shrnuje dosažené výsledky a nastiňuje možnosti dalšího vývoje.

# RGBD kamery

RGBD kamery poskytují kromě klasického barevného obrazu s RGB kanály i údaj o hloubce pro každý pixel. Typicky se skládají ze tří zařízení - RGB kamery, IR kamery a IR projektoru. RGB kamera poskytuje barevný obraz a dvojice IR kamera a projektor se stará o výpočet hloubky. Kromě obrazových senzorů tyto kamery často obsahují i další zařízení jako např. pole mikrofónů, inerciální jednotku anebo motorizovaný stojan.

První komerčně dostupnou RGBD kamerou byl Kinect od společnosti Microsoft. Původně byla tato kamera určena jako periferie ke konzoli Xbox 360 ale po přidání možnosti připojení k PC a poskytnutí SDK si našla široké využití i v jiných oblastech, zejména pak v robotice. S nástupem nové generace konzolí přišla i vylepšená verze Kinect 2 určená pro konzoli Xbox One, přičemž možnost připojení k PC pomocí adaptéru byla zachována. Už od první verze Kinectu poskytuje Microsoft SDK, které kromě přístupu k surovým datům (barevný a hloubkový obraz) umožňuje i sledování uživatele. Výsledkem tohoto sledování je kostra složená z kloubů, kde každý kloub má svou pozici a orientaci v prostoru. Celková kostra pak určuje pozici a pózu uživatele před kamerou. Výhodou Kinectu je poměrně velká uživatelská základna, jednoduchost použití a hlavně možnost sledování uživatele, nevýhodou je nyní horší dostupnost z důvodů ukončení jeho výroby. Jako alternativu k tomuto zařízení sám Microsoft na svých stránkách doporučuje konkurenční řešení Intel RealSense.

V rámci technologií Intel RealSense jsou nabízeny dvě samostatné kamery D415 a D435. Tyto kamery jsou svými vlastnostmi srovnatelné s Kinectem 2. Výhodou těchto kamer jsou menší rozměry a lepší dostupnost. Nevýhodou je absence SDK umožňující sledování uživatele. To je možné doplnit s využitím knihoven třetích stran jako např. OpenCV popř. NuiTrack. Kromě samostatných kamer nabízí Intel v rámci technologie RealSense i vestavitelné kamery SR300 a T260.

Další méně známé RGBD kamery jsou například Asus Xtion, Orbbec a LIPS. Všechny tyto kamery mohou posloužit jako alternativa jak ke Kinectu tak i k technologiím Intel RealSense. Nevýhodou u většiny z nich je menší uživatelská základna, horší parametry a částečná nebo i úplná absence možnosti sledování uživatele v rámci SDK.

Kromě RGBD kamer lze pro účely projektu využít i pasivní stereo kamery popř. klasické RGB kamery. Pro tyto kamery by se však sledování uživatele muselo vyvíjet zvlášť, protože často nelze použít ani řešení třetích stran, které spoléhají na informaci o hloubce, která u těchto kamer chybí. U stereo kamer lze informaci o hloubce získat. Předpokladem pro vytvoření hloubkové mapy stejné kvality jako u RGBD kamer je však vhodná scéna s množstvím detailu. Řešení pomocí těchto zařízení by bylo určitě zajímavé, hlavně z hlediska jejich jednoduché dostupnosti a velké rozšířenosti. Protože doplnění sledování uživatele by zabralo příliš mnoho času, jsou tyto zařízení uvedeny spíše jako možnost pro budoucí vývoj.

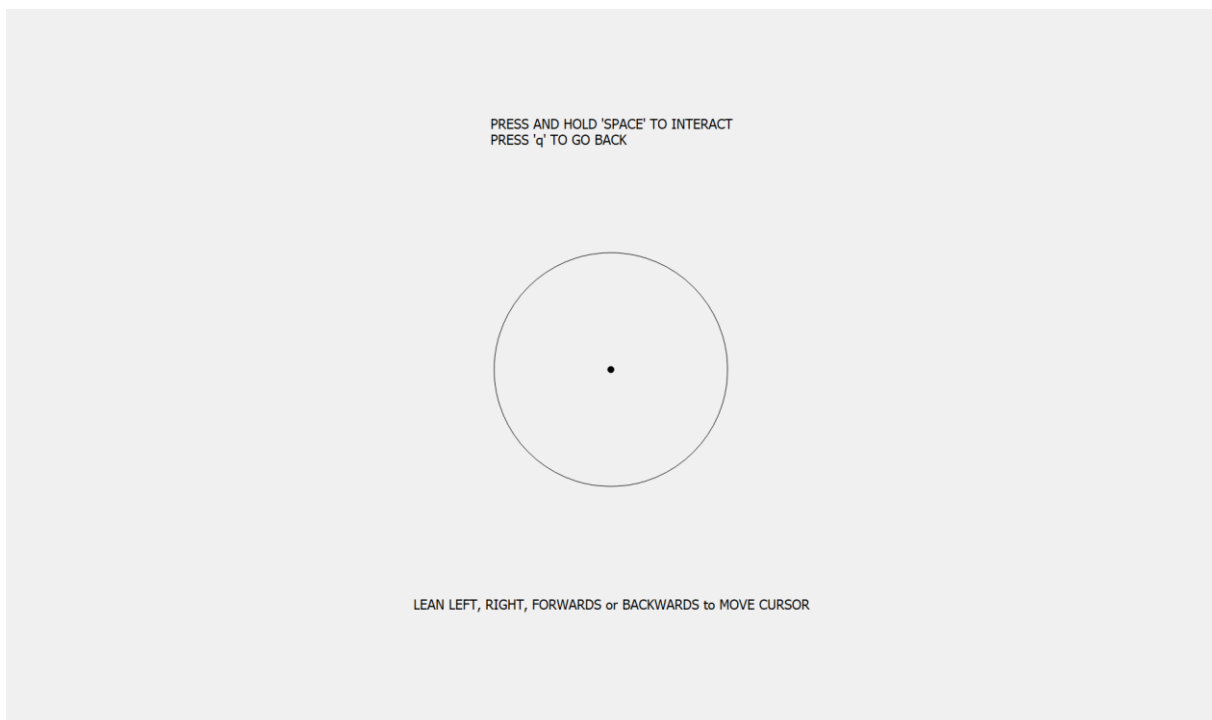
Cílem smluvního výzkumu je hlavně vyzkoušet vhodnost ovládní založeném na sledování uživatele. Proto je nejvhodnější použít RGBD kameru s integrovaným sledováním, což umožní rychlé vytvoření prototypu k testování. Jako nejlepší možnost se tedy jeví Kinect 2 od společnosti Microsoft, který jako jediný má sledování uživatele zabudované přímo v SDK od výrobce.

# Technické řešení

Tato kapitola popisuje technické řešení tří navrhovaných technik pro ovládání mikroskopu. Tyto techniky jsou postupně rozebrány a ke každé je uvedeno, jaké informace o uživateli jsou využity. Pro každou techniku je vytvořeno demo v jazyce C++ s využitím knihovny Qt ve verzi 5.5.1 a Kinect for Windows SDK 2.0. Všechny tři demo jsou integrovány do jedné aplikace, která umožňuje mezi těmito demy přepínat. Ovládání jednotlivých dem je založeno na sledování polohy a pózy uživateli hlavy pomocí senzoru Kinect 2, který je umístěn na monitoru a namířen na uživatele. Součástí aplikace je i napojení na API pro ovládání mikroskopu dodaném firmou TESCAN. Přes toto API jsou zasílány zpracované informace o poloze a póze uživateli hlavy, které jsou potom využity v simulátoru.

## Řízení dvojrozměrné funkce

Toto demo umožňuje uživateli specifikovat 2D souřadnice pomocí pohybu hlavy. Po detekci uživatele před kamerou je mu umožněno pomocí stisku klávesy mezerník nastavit výchozí pozici hlavy. Po uvolnění klávesy jsou vypočítávány ofsety aktuální polohy hlavy od nastavené výchozí polohy. Opětovným stiskem klávesy je možné nastavit novou výchozí pozici. Uživatel pomocí naklánění vpřed/vzad a vlevo/vpravo může nastavit libovolnou 2D souřadnici anebo 2D hodnotu. Ofsety jsou udávány v m. Vizualizaci této interakce znázorňuje obrázek 1. Obrázek ukazuje, že kromě nápovědy je uživateli zobrazen kruh se středovým bodem, jehož polohu může uživatel určit pomocí pohybu hlavy.

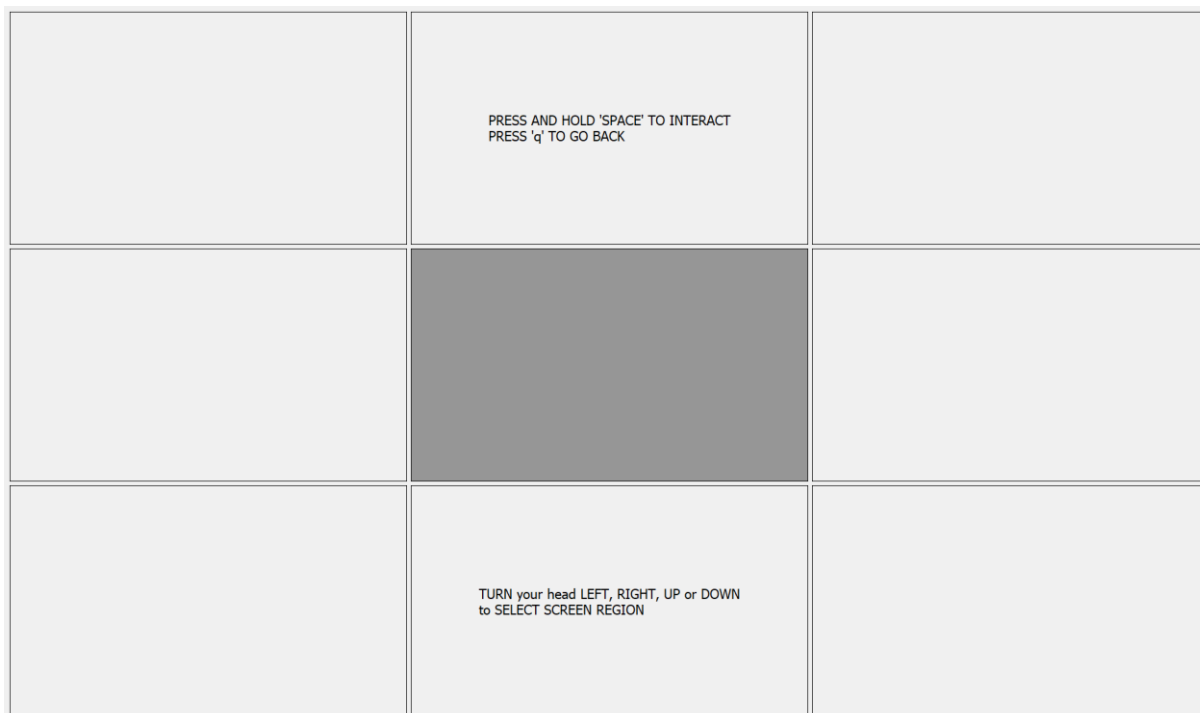


Obrázek 1: Grafické uživatelské rozhraní pro první demo - Řízení dvojrozměrné funkce.

## Automatické přepínání kontextu

Toto demo umožňuje uživateli přepnout kontext okna pouze pomocí pohledu bez nutnosti použít myš. Po detekci uživatele před kamerou je mu umožněno pomocí stisku klávesy mezerník nastavit výchozí pózu hlavy. Po uvolnění klávesy jsou vypočítávány ofsety aktuálního natočení hlavy od nastavené výchozí pózy. Opětovným stiskem klávesy je možné nastavit novou výchozí pózu. Uživatel pomocí natáčení hlavy nahoru/dolů a vlevo/vpravo může vybrat některé ze zobrazených oken a přepnout tak kontext. Ofsety jsou udávány ve stupních.

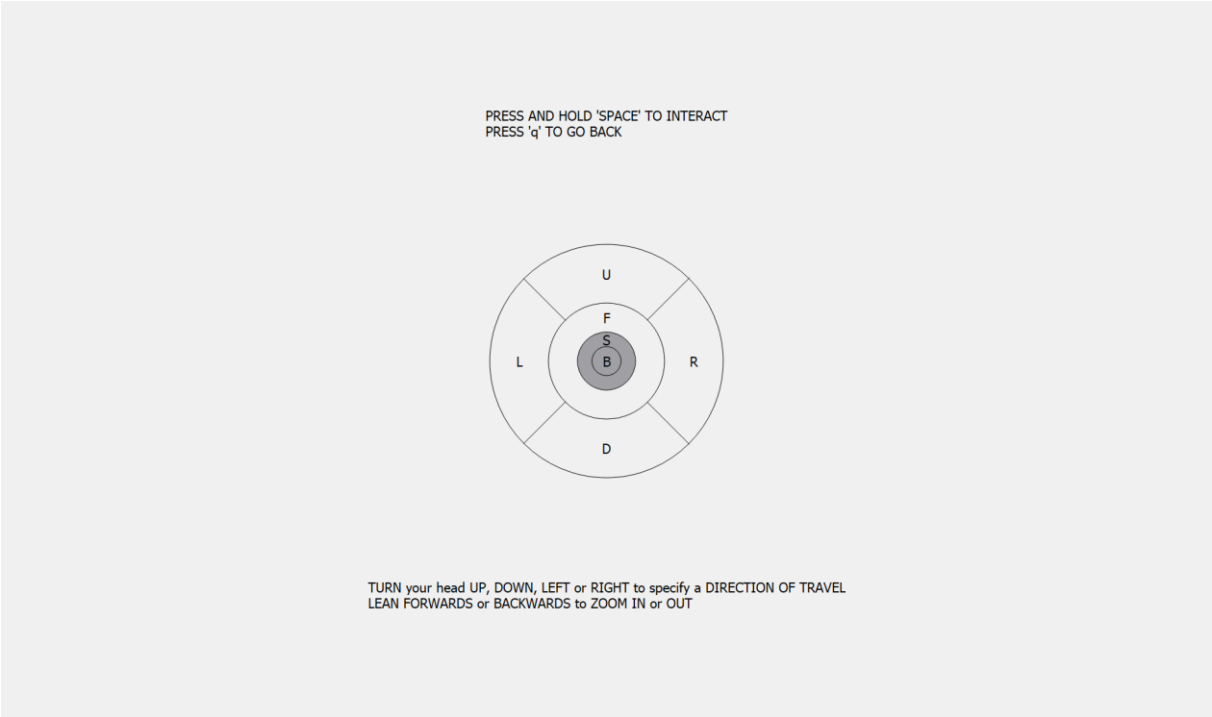
Vizualizaci této interakce znázorňuje obrázek 2. Obrázek ukazuje, že kromě nápovědy je uživateli zobrazena mřížka oken, ze které může pomocí pohledu vybrat požadované okno.



Obrázek 2: Grafické uživatelské rozhraní pro druhé demo - Automatické přepínání kontextu.

### **Navigace v rozlehlém 2D obraze**

Toto demo umožňuje uživateli pohybovat se v rozlehlém 2D obraze pomocí pohybu a natočení hlavy. Toto demo kombinuje techniky použité u předchozích dvou dem. Po detekci uživatele před kamerou je mu umožněno pomocí stisku klávesy mezerník nastavit výchozí polohu a pózu hlavy. Po uvolnění klávesy jsou vypočítávány offsety aktuální polohy a aktuálního natočení hlavy od nastavené výchozí polohy a pózy. Opětovným stiskem klávesy je možné nastavit novou výchozí polohu a pózu. Uživatel pomocí natáčení hlavy nahoru/dolů a vlevo/vpravo může vybrat odpovídající směr pohybu a pomocí naklánění vpřed a vzad přiblížit nebo oddálit obraz. Vizualizaci této interakce znázorňuje obrázek 3. Obrázek ukazuje, že kromě nápovědy je uživateli zobrazen navigační kruh, ve kterém se vyznačí aktuální směr pohybu.



Obrázek 3: Grafické uživatelské rozhraní pro třetí demo - Navigace v rozlehlém obraze 2D.

# Vyhodnocení

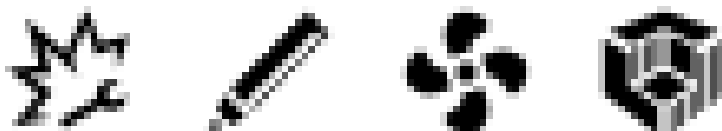
Tato kapitola popisuje vyhodnocení dvou implementovaných technik. Jedná se o techniky pro automatické přepínání kontextu a navigaci v rozlehlém 2D obraze. Testovací sestava se skládala z PC s připojeným Kinectem 2, který byl umístěn na monitoru. Na PC běžel simulátor mikroskopu, který komunikoval s vytvořenou aplikací běžící na pozadí. Vytvořená aplikace zpracovávala data ze senzoru a skrze domluvené rozhraní je zasílala do simulátoru. Uživatelé pracovali převážně se simulátorem, vytvořená aplikace byla pro ně transparentní. Pro každou testovanou techniku byla vytvořena umělá modelová úloha, která simulovala reálnou úlohu při práci s mikroskopem. Tyto modelové úlohy jsou popsány dále.

## Automatické přepínání kontextu

V této modelové úloze má uživatel za úkol nastavit jas a kontrast detektorů. Obrazovka je rozdělena na čtyři okna. Kde tři okna reprezentují jednotlivé barevné složky RGB modelu a čtvrté okno výsledek jejich složení. Uživatel má za úkol nastavit kontrast a jas jednotlivých složek, aby ve čtvrtém okně dosáhl požadovaného výsledku. Nastavování kontrastu a jasu provádí uživatel postupně dvěma způsoby. Pomocí myši a klávesnice, kde nejprve pomocí myši musí vybrat požadované okno (RGB složku), poté v kontextové nabídce kliknout na ikonku pro nastavování jasu a kontrastu a nakonec nastavit požadované hodnoty buď přímým zadáním v Padu anebo pomocí klávesových zkratk Alt + kurzorové šipky. Pomocí vytvořeného rozhraní pro automatické přepínání kontextu popsaného v předchozí kapitole. Oproti myši umožňuje vytvořené rozhraní vybrat okno i režim nastavování jasu a kontrastu zároveň bez nutnosti klikat na ikonku v kontextové nabídce. Zadání požadovaných hodnot se provádí stejným způsobem jako v případě myši a klávesnice. Tuto úlohu uživatelé opakovali pouze jednou pro každé rozhraní. Úloha končila, když uživatel nastavil požadovanou barvu, přičemž o tom, jestli nastavená barva odpovídá zadání rozhodoval sám uživatel.

## Navigace v rozlehlém 2D obraze

V této modelové úloze má uživatel za úkol najít drobný předmět v rozlehlém obraze. Uživateli je předložen obraz o velikosti 12000 x 12000 px do kterého je vložen menší obrázek o velikosti 20 x 20 px (viz obrázek 4). Vložený obrázek se značně liší od zbytku obrázku, a pokud na něj uživatel narazí, tak si bude naprosto jistý, že našel požadovaný objekt. Navigaci v rozlehlém obraze uživatel provádí postupně dvěma způsoby. Nejprve pomocí myši, kde stiskem levého tlačítka a tažením je možné obraz posouvat a pomocí kolečka myši je možné zoomovat na pozici kurzoru a poté pomocí vytvořeného rozhraní pro navigaci v obraze popsaného v předchozí kapitole. Celkově byly připraveny 4 obrazy, do kterých byly vloženy 4 různé objekty na 4 různé místa. Dva z těchto obrazů byly použity pro první způsob interakce a zbylé dva pro druhý. Tuto úlohu tedy uživatelé opakovali dvakrát pro každé rozhraní. Prohledávání obrazu bylo ukončeno po nalezení objektu anebo po uplynutí časového limitu 5 minut.



Obrázek 4: Objekty vložené do prohledávaného obrazu.

Testování se zúčastnili celkem 4 studenti. Jednalo se o technicky zdatné studenty v posledních ročnících bakalářského (3 studenti) a magisterského studia (1 student). Všichni uživatelé postupně vyzkoušeli obě modelové úlohy s existujícím i nově navrženým rozhraním.



Po každé úloze bylo uživateli položeno několik otázek zaměřených na pohodlnost, pochopitelnost, jednoduchost a na problémy při použití testovaných rozhraní. Sezení jednoho uživatele trvalo asi 45 minut.

Vytvořené rozhraní pro automatické přepínání kontextu hodnotili uživatelé pozitivně. Připadalo jim pohodlné, intuitivní a jednoduché na použití. Jako námět na zlepšení uváděli lepší zvýraznění vybraného okna. Oproti použití myši uváděli jako výhodu, že není potřeba zvedat ruce z klávesnice a přehmatávat na myš. Při použití myši jim vadila hlavně nutnost klikat na ikonku v kontextové nabídce. Ani s jedním rozhraním neměli uživatelé problém úlohu dokončit.

U vytvořeného rozhraní pro navigaci v obraze uživatelé uváděli, že je pro ně jednoduché na pochopení a intuitivní. Rozhraní jim připadalo příliš citlivé, zároveň však uváděli, že si na ně lze po troše tréninku zvyknout. Problém jim činilo hlavně zoomování, na které se museli více soustředit a pokud to šlo tak se mu raději vyhýbali. U zoomování měli hlavně obavu z toho, že se po zazoomování na objekt a jeho prohlédnutí nedokáží vrátit do výchozí pozice, ze které zoomování započali a ze které chtěli dále pokračovat v prohledávání obrazu. S pohybem do stran problémy nebyli a na rozdíl od zoomu jim připadal pohodlný. Jediným problémem, který nastával při pohybu do stran byla chybějící zpětná vazba při dojetí na okraj obrazu. Po dojetí na okraj totiž pohyb ustal a uživatelé si často nebyli jisti, jestli systém zaznamenává jejich pozici hlavy. Jako námět na vylepšení uvedli přesunutí zoomování na myš nebo klávesnici a změnu lineární interpolace rychlosti pohybu za exponenciální. U této úlohy uživatelé spíše preferovali interakci pomocí myši, na kterou jsou zvyklí a se kterou neměli při zoomování problémy a u které nenastávala nejistota při dojetí na okraj. Z celkem 8 pokusů o nalezení objektu u každého rozhraní byli uživatelé úspěšní v 5 případech při použití myši a ve 3 případech při použití Kinectu.

# Závěr

Tato zpráva popisuje návrh, implementaci a vyhodnocení experimentálního uživatelského rozhraní pro interakci s elektronovým mikroskopem založeném na sledování pohybu hlavy uživatele. Toto rozhraní bylo připraveno v rámci smluvního výzkumu s firmou TESCAN. Popis zadání a úvod do problematiky je rozepsán v první kapitole. Pro sledování pohybu uživatele byla použita RGBD kamera Kinect 2, která je představena společně s možnými alternativami v druhé kapitole. Zaznamenaný pohyb byl interpretován vytvořenou aplikací a přes domluvené rozhraní posílán do simulátoru elektronového mikroskopu. Celkem byly navrženy a implementovány 3 techniky a to pro řízení dvojrozměrné funkce, automatické přepínání kontextu a navigaci v rozlehlém 2D obraze. Technické řešení těchto technik je rozebráno ve třetí kapitole. Vyhodnoceny byly poslední dvě zmíněné techniky.

Vyhodnocení probíhalo formou uživatelských testů, kterého se zúčastnili 4 studenti. Po provedení zadaných modelových úloh s pomocí existujícího i nově vytvořeného rozhraní byla od uživatelů sesbírána zpětná vazba. Výsledky testování jsou shrnuty ve čtvrté kapitole společně s možnými náměty na vylepšení testovaných rozhraní.