

Virtualna okruženja

Laboratorijske vježbe

Vježba 5

Virtualni ljudi

FER – ZTE – Igor S. Pandžić
Suradnja na pripremi vježbe:
Aleksandra Čereković
Josip Iveković

Uvod

Virtualni ljudi ili virtualni likovi (engl. *virtual characters*) su su simulacija ljudi na računalu pri kojoj virtualni model izgleda i ponaša se manje ili više slično stvarnom čovjeku. U osnovi, simulacija ljudi nije bitno različita od simulacije raznih virtualnih predmeta –definira se geometrija, materijali i animacija. Danas su virtualni ljudi toliko napredovali da se grafički modeli ne razlikuju od stvarnih ljudi. Ipak, simulacija ponašanja bitno zaostaje za modelima. Pažljivo izrađeni likovi se često ponašaju poput robota, nemaju emocija, „ne razumiju“ i „ne čuju“ neke podražaje iz okoline, imaju sintetičke glasove i ne čine se odveć pametni.

Cilj ove laboratorijske vježbe je upoznati se s virtualnim ljudima i animacijom preko animacije lica i tijela kreiranih na nekoliko različitih načina.

1. Alati potrebni za izvođenje vježbe

Za izvođenje vježbe potrebno je imati VisageSDK (nalazi se na stranicama predmeta) i web kameru.

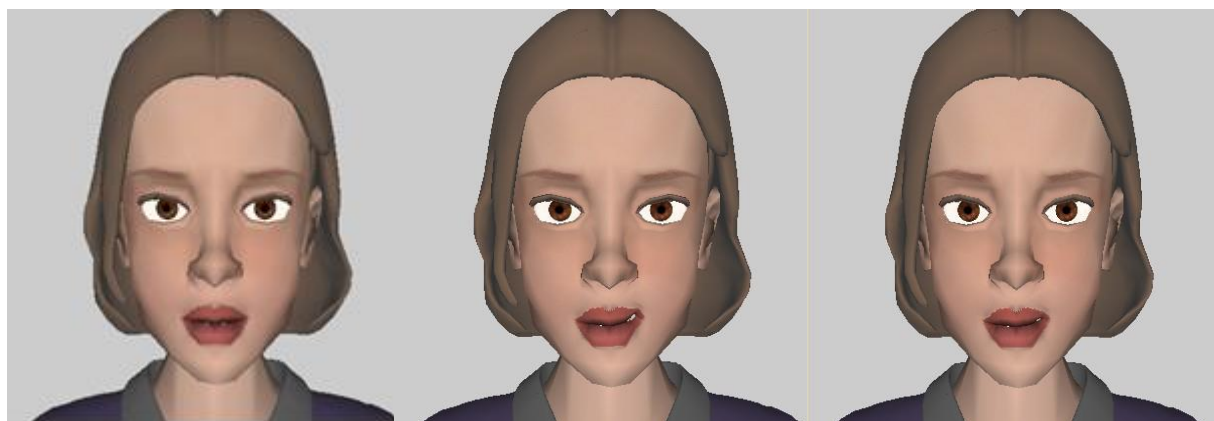
2. Teorijska podloga

Animacija ljudi ima dosta specifičnosti te se zbog svoje složenosti promatra kao zasebno područje animacije. Štoviše, zbog svoje složenosti sama animacija ljudi dijeli se na animaciju lica i tijela te animaciju specifičnih dijelova ljudskog tijela, poput animacije kose, odjeće, suza, bora i sl.

2.1 Animacija lica

Animaciju lica možemo podijeliti u dva područja: animacija na niskoj i na visokoj razini. Premda se postojeći sustavi ne dijele strogo na ovaj način, tj. često su integrirani, daleko je jednostavnije promatrati animaciju lica podijeljenu na ove dvije funkcionalne cjeline.

Animacija na niskoj razini podrazumijeva pokretanje geometrije lica (pomaci usana, obrva, očiju itd.) pomoću nekog skupa parametara. Parametri su obično relativno niske razine, npr. parametar za podizanje vanjskog ruba lijeve obrve, parametar za istezanje lijevog kuta usana i sl. Važno je imati takav skup parametara kojima se lice može deformirati u bilo koji izraz. Jedan primjer animacije na niskoj razini je interpolacija. Slika 1 prikazuje osnovnu ideju interpolacije. Pretpostavimo da imamo model lica u neutralnom položaju i lice s otvorenim ustima. Jednostavnom interpolacijom položaja vrhova u modelu možemo dobiti sve međupoložaje (Slika 1). Time smo praktično implementirali parametar za otvaranje usta. Model s otvorenim ustima može se pripremiti ručno ili nekom drugom metodom. Ovaj model, koji služi kao krajnja točka interpolacije, naziva se interpolacijski cilj (engl. *morph target*).



(a)

(b)

(c)

Slika 1. Primjer interpolacije a) neutralan položaj, b) interpolacijski cilj i c) međupoložaj

Animacija na visokoj razini koristi skup parametara niske razine za stvaranje čitavih animacijskih sekvenci. Animacijska sekvenca je vremenski slijed animacijskih parametara niske razine. Animacija na visokoj razini uključuje raznovrsne metode stvaranja ovakvih sekvenci, i može se promatrati potpuno odvojeno od animacije na niskoj razini, tj. načina na koji se animacijski parametri pretvaraju u stvarne pokrete lica. Postoji izuzetno širok spektar metoda za animaciju, a jedne od njih su animacija iz teksta, animacija pomoću glasa i animacija primjerom.

2.1.1 Animacija iz teksta

Pri animaciji lica vrlo često je potrebno simulirati govor, dakle postići da se usne modela lica pomiču sinkronizirano sa zvukom govora. Jedan od načina da se to postigne je animacija iz teksta. Ovo je potpuno automatska metoda. Pritom se koristi sinteza govora (engl. *Speech synthesis, text-to-speech*), odnosno sinteza zvuka iz teksta. Pri sintezi proizvode se i animacijski pokreti lica (usana), te se na taj način potpuno automatski iz običnog teksta dobiva animacijska sekvenca lica koje govori. Ovo se naziva vizualna sinteza govora (engl. *Visual text-to-speech, VTTS*).

Sinteza govora je danas već standardno prisutna na mnogim osobnim računalima. Široko rasprostranjeni standard SAPI (*Speech API*) omogućuje aplikacijama da na jednostavan i standardiziran način koriste standardne pakete za sintezu govora raznih proizvođača, čime postaje prilično jednostavno podržavati razne jezike. Sinteza govora vodećih proizvođača kao što je AT&T približava se kvaliteti ljudskog glasa. Sinteza govora uz zvuk proizvodi i seriju fonema s odgovarajućim vremenima trajanja (fonem je osnovna jedinica govora, jedan glas). Na osnovu ovih fonema i njihovih trajanja, vrlo se jednostavno može postići osnovna animacija usana sinkroniziranih s govorom, na taj način da se za svaki fonem odabere jedan oblik usana, te se oblici mijenjaju u skladu sa sintetiziranim fonemima. Pošto čitave grupe fonema dijele oblik usana koji izvana izgleda jednako (npr. p, b, m), koriste se tzv. vizemi. Vizem je karakteristični oblik usana za izgovor fonema. Svi fonemi za koje je vanjski oblik usana jednak preslikavaju se na isti vizem. Sa 10 – 15 vizema moguće je napraviti razmjerno dobru animaciju. Na primjer, standard MPEG-4 definira 14 vizema. Navedeni pristup ima i poznate nedostatke. Na primjer, potpuno naivni pristup izravnog preslikavanja fonema na vizeme, te naglog mijenjanja oblika usana na vremenskoj granici među fonemima daje prilično loše

rezultate zbog pojave koja se naziva koartikulacija. Naime, u trenutku izgovaranja pojedinog glasa, usne su još pod utjecajem prethodno izgovorenog glasa, a istovremeno se već pripremaju za slijedeći. Pokazano je da u bilo kojem trenutku oblik usana tokom govora može ovisiti o pet susjednih glasova. Da bi se dobili dobri rezultati animacije lica iz teksta, potrebno je uključiti model koartikulacije. Dodatni nedostatak ovakvih sustava je to što proizvode samo pokrete usana, dok u stvarnosti govornici proizvode i razne druge pokrete lica, geste ili emocije. Ovo se može donekle riješiti dodavanjem proizvoljnih pokreta lica ili pokreta snimljenih s prirodnog govornika, a napredniji sustavi pokušavaju iz leksičke analize teksta i statističkih modela ponašanja pri govoru automatski proizvesti što prirodnije pokrete. Vizualna sinteza govora, kao potpuno automatska metoda koja uz to radi u stvarnom vremenu, vrlo je podobna za interaktivne primjene, a može se koristiti i kao osnova za dodatnu ručnu doradu.

2.1.2 Animacija pomoću glasa

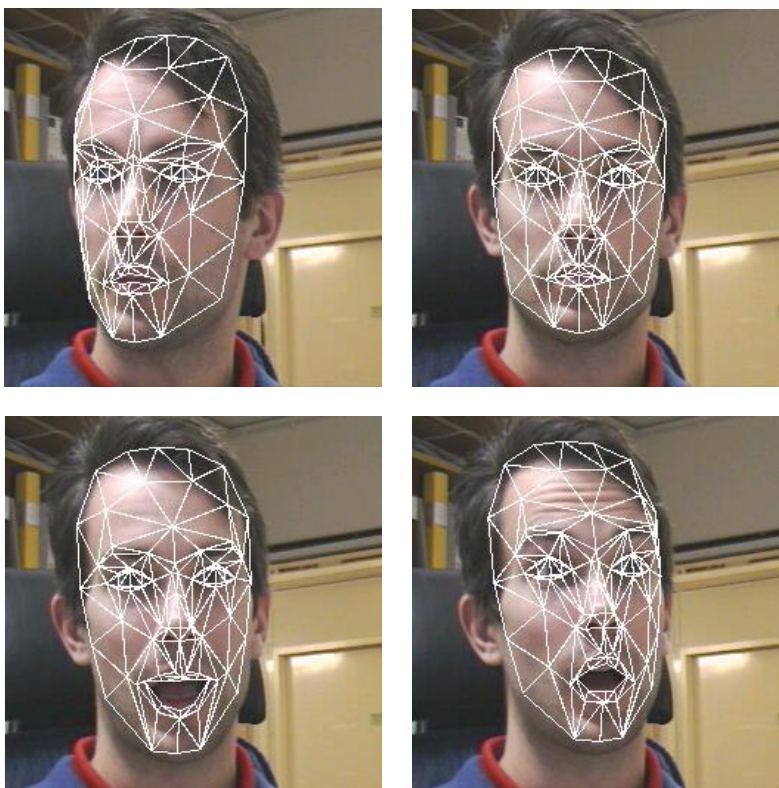
Drugi način animiranja govora je automatska sinkronizacija usana na postojeći glas (engl. *lip sync*). Zasniva se na analizi zvuka, kojom se nastoje dobiti parametri iz kojih se može zaključiti odgovarajući položaj usana u svakom trenutku. Najprimitivniji primjer je jednostavno praćenje glasnoće signala zvuka i otvaranje/zatvaranje usta proporcionalno s glasnoćom. Naravno, ovakvom naivnom metodom dobivaju se pokreti usana koji tek nezgrapno prate govor, otprilike na razini nekih starijih crtanih filmova. Mogu se koristiti metode za prepoznavanje govora, koje su također gotovo standardno prisutne na današnjim računalima. Ove metode prepoznaju čitave riječi i rečenice, a na osnovu tog prepoznavanja mogu proizvesti vremenski slijed fonema prema kojem se može vršiti animacija. No, za animaciju lica nas ne zanimaju čitave riječi, te je stoga ova metoda zapravo kompleksnija nego što je potrebno. Uz to, ukoliko se riječ ne prepozna, ne dobiva se nikakav rezultat te animacija za taj dio jednostavno nedostaje. Napokon, svaka aplikacija za raspoznavanje govora napravljena je za jedan jezik. Za animaciju pomoću glasa bilo bi idealno da bude univerzalna, tj. da radi za bilo koji jezik. Iz tih razloga bolje je koristiti metode analize zvuka na nižoj razini, koje ne pokušavaju raspoznavati čitave riječi, pa čak ni foneme, nego u kratkim vremenskim segmentima analizom raščlanjuju zvuk na skup parametara iz kojih se zaključuje oblik usana, tj. vizem. Jedan primjer takve metode je linearna predikcija, metoda poznata iz kodiranja zvuka kod koje se segment zvuka (obično reda veličine 30 ms) raščlanjuje na parametre govornog modela, a iz tog skupa parametara zaključuje se koji oblik usana (vizem) je trenutno aktivan.

2.1.3 Animacija primjerom

Animacija primjerom (engl. *performance animation*) je animacija virtualnog lica koja se dobiva upravljanjem, odnosno kopiranjem pokreta stvarnog lica na virtualno lice pomoću upravljačkog sučelja. Upravljanje stvarnim - ljudskim licem je najprirodnije sučelje za upravljanje virtualnim licem i proizvodnju animacije, no na žalost nije ga jednostavno implementirati.

Prilično dobri rezultati dobivaju se sa sučeljem u kojem se markeri lijepe na ljudsko lice i zatim prate jednom standardnom video kamerom. Pritom se algoritmi analize slike koriste za slijeđenje markera u slici. No, lijepljenje markera na lice vrlo je nepraktično i neudobno, a za pojedine primjene i potpuno neprihvatljivo. Stoga se već godinama intenzivno radi na algoritmima koji bi slijedili pokrete lica u običnoj video slici, bez uporabe markera na licu. Taj postupak se naziva praćenje karakterističnih točaka lica (engl. *Facial Feature Tracking* ili *Facial Motion Capture*) ili skraćeno praćenje lica (engl. *Face Tracking*). Ovo je izuzetno

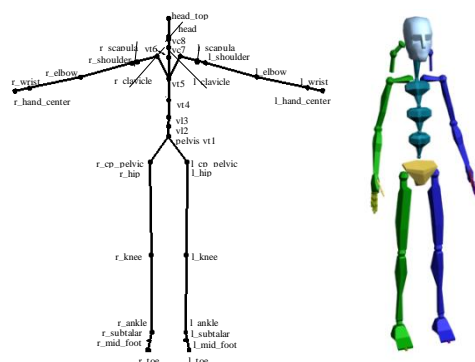
složen problem čak i u idealnim uvjetima, a postaje još složeniji kada se uzmu u obzir razlike u kvaliteti osvjetljenja, boji kože, mogućnost da korisnik ima bradu ili nosi naočale, itd. U posljednjih pet godina pojavila su se rješenja koja rade relativno stabilno i ugrađuju se u komercijalne proizvode (Slika 2).



Slika 2. Sljedeđenje iz čiste video slike

2.2 Animacija tijela

Animacija tijela se gotovo uvijek zasniva na modelu kostura koji je „smješten“ unutar modela tijela. Tijelo se pokreće pokretima elemenata kostura, a to gibanje se prenosi na vidljivu površinu modela, tj. kožu. Kostur je organiziran u hijerarhijsku strukturu (Slika 3) i nije grafički predstavljen, tj. ne crta se nego služi samo kao matematički model za upravljanje tijelom. Pokreće se skupom parametara kojim mijenjanjem animiramo tijelo. Ti parametri su stupnjevi slobode gibanja (SSG, engl. *Degrees of Freedom*, DOF). Najčešće se radi o rotacijama zglobova u kosturu, što je donekle slično upravljanju u robotici.



Slika 3. Modela kostura u MPEG-4 standardu (75 SSG) i model iz 3ds Max studija

Veza između kostura i kože izuzetno je važna za kvalitetu krajnje animacije, jer ona određuje završni izgled tijela u svakom trenutku. Najjednostavniji način izvođenja ove veze je sastavljanje tijela od krutih segmenata koji su izravnom, krutom vezom spojeni na model kostura, tj. na pojedine kosti. Bolje metode uključuju međuslojeve između kostura i kože.

Slično kao i animaciju lica, u funkcionalnom smislu, animaciju tijela možemo podijeliti na animaciju niske i animaciju visoke razine. Animacija niske razine bavi se parametrima koji pokreću kosti i utječu na izgled animacije, a to su veze između kostiju, problem veze kostura i kože, te deformacija kože na tijelu i ruci putem simulacije. Poput animacije lica, i animacija tijela na visokoj razini koristi skup parametara niske razine za stvaranje čitavih animacijskih sekvenci. To mogu biti primjeri snimljenih animacija, npr. animacije modelirane u 3ds max-u ili animacije dobivene slijeđenjem (engl. *motion capture*). Također, animacije visoke razine mogu se modelirati proceduralno, odnosno matematičkim aproksimacijama vanjskih sila ili inverznom kinematikom, čime se simuliraju osnovni zakoni fizike. Posljednji i najkompleksniji slučaj animacije visoke razine je hibridna animacija, odnosno animacija iz primjera koja objedinjuje posljednja dva principa. Ovaj princip se odnosi na obradu snimljenih podataka sa ciljem dobivanja novih realističnih animacija. Primjer ove vrste animacije je miješanje pokreta (engl. *motion blending*), postupak u kojem se težinski miješaju dvije ili više snimljenih animacija kako bi se dobile nove animacije.

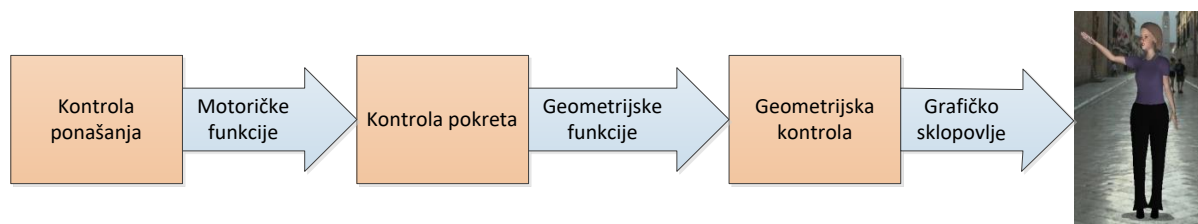
2.3 Standardi za virtualne ljude

Postoji mnoštvo sustava i metoda za sve do sada nabrojane segmente simulacije ljudi, no svaki sustav koristi vlastiti skup parametara za izražavanje, zapis i prijenos izgleda i animacije. Naravno da takva situacija otežava razmjenu znanja i iskustava, te koči komercijalnu uporabu, razvoj novih primjena i otvaranje novih tržišta za te primjene. Nakon više od dva desetljeća rada na području virtualnih ljudi, tijekom devedesetih godina prošlog stoljeća dolazi do interesa za standardizaciju. Tijekom devedesetih godina skupina stručnjaka iz industrijskih i akademskih istraživačkih krugova radi na izradi standarda za virtualne ljude u okviru međunarodnih standardizacijskih tijela, prvenstveno Međunarodne Organizacije za Standarde ISO. Pod okriljem ISO-a, a u sklopu širih standarda MPEG-4 i VRML, nastaju specifikacije za virtualne ljude. U sklopu standarda MPEG-4 nastaje specifikacija za opis oblika i animacije ljudskog tijela i lica pod nazivom Face and Body Animation (FBA). U sklopu VRML-a postoji radna skupina pod nazivom H-ANIM (skraćeno od engl. *human animation*, animacija ljudi) koja je donijela specifikaciju za animaciju tijela unutar postojećeg VRML standarda. Ova specifikacija je nastala u suradnji sa skupinom MPEG-4 FBA, te je zapravo implementacija dijela MPEG-4 specifikacije unutar VRML-a.

2.4 Koncept animacijskog sustava za virtualne ljude

Postojeće animacijske tehnike (tijelo, lice, odjeća, kosa...) integriraju se u animacijski sustav. On objedinjuje mehanizme koji na najvišoj razini upravljaju redoslijedom izvođenja pojedinih animacija. Osim upravljivosti, animacijski sustav mora omogućiti uvjerljivost i realističnost animacija koje izvodi na virtualnom liku.

Pod animacijskim sustavom za virtualne ljude podrazumijevamo sustav koji pomoću ulaznih parametara proizvodi animacijske sekvence koje vizualiziraju, odnosno predstavljaju ponašanje virtualnog čovjeka. Slika 4 opisuje koncept animacijskog sustava.



Slika 4. Shema animacijskog sustava za virtualne ljude

Unutar animacijskog sustava postoje tri razine kontrole:

- Najviša razina: kontrola ponašanja, sinkronizacija
- Srednja razina: kontrola pokreta, motorika- pomicanje dijelova tijela
- Niska razina: geometrijska kontrola, pomicanje elemenata 3D modela

Najviša razina (kontrola ponašanja) odvija se pomoću skupa parametara visoke razine apstrakcije koji opisuju ponašanje virtualnog čovjeka. Najčešće, riječ je o primitivnim ponašanjima, poput gesti ruku, kretnje u prostoru, pomicanja glave, gesti lica i sl. Korištenje tih parametara izvodi se u određenom vremenskom slijedu. Stoga, potrebno je omogućiti međusobnu sinkronizaciju parametara kako bi se njihovim izvođenjem dobila različita ponašanja. Primjer je ponašanje u kojem se gesta kimanja glave sinkronizira sa rečenicom „Da, vidjela sam ga prošli tjedan“. Ako omogućimo međusobnu sinkronizaciju parametara, gestu kimanja možemo izvesti prije te rečenice, nakon nje, ili je izvesti prilikom izgovora pojedine riječi. Na taj način dobivamo različita ponašanja. Najčešće, skupovi ovih parametara i principi sinkronizacije opisuju se skriptnim jezikom. Primjer je jezik Behavior Markup Language (engl. BML).

Motoričke funkcije preslikavaju ove parametre na animacije visoke razine, odnosno, odlučuju koje će se od tehnika animacija visoke razine koristiti za animaciju pojedinog primitivnog ponašanja. Primjeri tehnika visoke razine za animaciju tijela su primjeri animacija, proceduralne animacije i hibridni pristup.

Srednja razina (kontrola pokreta) obuhvaća parametre koji direktno kontroliraju pokrete modela, a to su stupnjevi slobode gibanja kostura (SSG) te parametri koji kontroliraju geometrijski model lica (npr. FACS ili MPEG-4 FBA).

Geometrijske funkcije i kontrola niske razine (geometrijska kontrola) odnose se na animaciju niske razine: pomicanje poligona i modela, parametrizaciju modela lica, pseudomišići i sl...

Na koncu, grafičko sklopovlje crta virtualni lik i animira ga uz pomoć niskorazinske kontrole. Prema najnovijim trenovima, cilj je da animacijski sustav bude neovisan o grafičkom procesoru.

2.5 Opis zadatka

Zadatak ove laboratorijske vježbe je upoznati se sa sustavom za praćenje i animaciju lica koje imamo na Zavodu.

3. Upute za rad

3.1 Visage SDK – Face Tracker

Visage tracker je softverski sustav za praćenje lica u stvarnom vremenu koji je djelomično razvijen i na Zavodu za telekomunikacije. Sustav je povezan sa softverom za animaciju lica čime se dobiva sustav koji animira virtualno lice u stvarnom vremenu, prateći direktno pokrete osobe pred kamerom.

Sustav za praćenje koristi Candide-3 model lica koji se sastoji od poligona, a koji se automatski postavlja na lice korisnika pri pokretanju kamere. Prije početka praćenja, univerzalni Candide-3 model se može prilagoditi obliku lica korisnika, odnosno modificirati, čime se dobivaju bolji rezultati praćenja. Ipak, u ovoj će se vježbi koristiti potpuno automatsko prilagođavanje lica korisniku (radi jednostavnosti). Za rad sa sustavom, potrebno je priključiti (i po potrebi instalirati) kameru na računalo.

Preuzmite i instalirajte VisageSDK sa stranica predmeta. Visage SDK sastoji se od nekoliko aplikacija koje se nalaze u direktoriju *visageSDK\bin64* u direktoriju u kojem je SDK instaliran. Za ovu vježbu bitan je *FacialAnimationUnity.exe*. Na stranicama kolegija dostupna je licenca koju je potrebno postaviti u direktorij u kojem se nalazi *FacialAnimationUnity.exe*.

3.1Zadatak

Vaš zadatak je da proučite aplikaciju za praćenje lica i da praćenjem vlastitog lica animirate virtualno lice. Pokrenite aplikaciju *FacialAnimationUnity.exe*. Ako niste preuzeli licencu, po pokretanju aplikacije pojavljuje se upozorenje te se aplikacija gasi sama od sebe nakon 60 sekundi korištenja, ali nakon toga je možete ponovno pokrenuti te je izrada vježbe zbog kratkih zadataka moguća i na taj način.

Nakon pokretanja aplikacije sustav će započeti sa praćenjem Vašeg lica. Maska se automatski postavlja na Vaše lice, koje mora biti u neutralnom položaju. Pratite lice i zaključite. Je li sustav stabilan na brzinu pomicanja glave? Što se događa pri rotaciji glave u različitim smjerovima? Prati li sustav pomicanje očnih kapaka i očiju? Pratite što se događa sa sustavom ako sklonite dio lica sa vidnog kruga kamere (napraviti sliku ekrana i priložiti u izvještaju). Kako se pokretanje odražava na animaciju? (2 slike ekrana, priložiti u izvještaju)

Odgovore na potonja pitanja i potrebne screenshotove priložite u izvještaju.

4. Predavanje rezultata vježbe

Rezultati vježbe se predaju zapakirani u PDF datoteku **VO-V5- Rezultati-*<ImePrezime>*** koja treba sadržavati izvještaj o izvođenju vježbe u kojem se nalazi opis postupka izrade cijele vježbe koji je zadan u zadatku za izradu.

Navedena arhiva treba biti predana korištenjem sustava *Moodle* koji je dostupan preko stranica predmeta.

Napomena: Rezultati se šalju isključivo preko gore navedene aplikacije. U slučaju problema, javiti se email-om na adresu vo@fer.hr. Sačuvajte kopiju poslanih rezultata.