

صلى الله عليه وسلم



دانشگاه سمنان

دانشکده مهندسی برق و کامپیوتر

پایان نامه کارشناسی ارشد مهندسی مکترونیک

شناسایی علائم زبان ایما و اشاره فارسی و تبدیل آن به نوشتار

دانشجو: علی اکبر نجف زاده

استاد راهنما: دکتر هادی سلطانی زاده

اسفند ۱۳۹۰

تقدیم بہ شور زندگی ام مادرم

تقدیم بہ شعور زندگی ام پدرم

تقدیم بہ تکیہ گاہ تنہائی بایم خواہرم

تقدیم بہ تمام زندگی ام، ہمسرم

و تقدیم بہ روح سعید و خلیل

تشکر و قدردانی

سپاس خداوند را که جانم از اوست. در راه تعالی و پیشرفت پدر و مادری مهربان و دلسوز و همسری فداکار همراهم ساخت. زبان از شکر این نعمات قاصر است. لیکن به رسم "من لم یشکر المخلوق لم یشکر الخالق" بر خود لازم می‌دانم در این مختصر، کمال امتنان و تشکر خود را تقدیم این عزیزان نمایم.

از تمامی دوستان عزیزم بالاخص جناب آقای مهندس رضا همتی که در این مدت در کنار آنها درس‌ها آموختم، تشکر می‌کنم. تشکر ویژه خود را به محضر سرکار خانم نگین ملک احمدی که همچون خواهری دلسوز در کنارم بودند، تقدیم می‌نمایم. در پایان نهایت تشکر و قدردانی را از استاد گرانقدر و وارسته جناب آقای دکتر سلطانی زاده که برای من سعادت‌ی بود در این مدت در محضرشان به کسب علم بپردازم، تقدیم می‌دارم.

چکیده

در این تحقیق، به شناسایی زبان ایما و اشاره فارسی^۱، به عنوان ابزاری ارتباطی برای تعامل انسان با کامپیوتر^۲، پرداخته می‌شود. زبان ایما و اشاره فارسی تقریباً دارای ۱۰۷۵ الفبا و کلمه رایج، شامل حرکات استاتیک و دینامیک می‌باشد که در این تحقیق بر روی الفبای استاتیک این زبان تمرکز شده است. در این تحقیق از هیچ‌گونه دستکش، پوشش و یا علامت‌گذارهای تصویری، برای اجرای حرکات و جمع‌آوری مجموعه داده، استفاده نشده است و سیستم با تصاویر دستان برهنه سرو کار دارد، که این امکان را به کاربر می‌دهد تا به صورت طبیعی با سیستم تعامل داشته باشد. برای شناسایی و تشخیص این علائم از روش شناسایی مبتنی بر استخراج ویژگی‌های محلی تغییرناپذیر با اندازه (SIFT^۳) استفاده شده است. سیستم پیاده‌سازی شده بر مبنای روش SIFT قادر به تشخیص ۳۷ الفبای ایما و اشاره فارسی با دقت تشخیص ۹۵/۱۴٪ می‌باشد. برای پیاده‌سازی الگوریتم‌ها از نرم‌افزار برنامه‌نویسی Delphi استفاده شده است.

کلید واژه

زبان ایما و اشاره فارسی، تعامل انسان با کامپیوتر، دستان برهنه، استخراج ویژگی‌های محلی،

Delphi، SIFT

^۱ Persian Sign Language (PSL)

^۲ Human-Computer Interaction (HCI)

^۳ Scale Invariant Feature Transform (SIFT)

فهرست مطالب

۱- مقدمه	۱
۲- بررسی مطالعات انجام شده	۴
۱-۲- مقدمه	۵
۱-۲- تحقیقات انجام شده در حوزه روش های پوشش محور	۶
۱-۱-۲- روش های مبتنی بر ابزارهای الکترومکانیکی	۶
۲-۱-۲- روش های مبتنی بر دستکش های رنگی	۸
۲-۲- تحقیقات انجام شده در حوزه روش های تصویر محور محض	۹
۳-۲- نرمالیزه کردن فواصل مرکز دست و مرز آن و استفاده از مدل های ANFIS	۱۲
۱-۳-۲- ساختار	۱۲
۲-۳-۲- استخراج ویژگی ها	۱۳
۳-۳-۲- تشخیص حرکت	۱۷
۴-۳-۲- نتایج	۱۸
۴-۲- استفاده از تبدیل هاف و شبکه های عصبی MLP برای تشخیص	۱۸
۱-۴-۲- نگاه کلی به سیستم	۱۹
۲-۴-۲- مرحله ی استخراج ویژگی	۲۰
۳-۴-۲- مرحله دسته بندی	۲۵
۴-۴-۲- نتایج	۲۷
۵-۲- استفاده از تبدیل موجی و شبکه های عصبی	۲۸
۱-۵-۲- ساختار سیستم ارائه شده	۲۸
۲-۵-۲- استخراج ویژگی ها	۲۹
۳-۵-۲- مرحله دسته بندی	۳۳
۴-۵-۲- آزمایشات و نتایج شبیه سازی	۳۵
۶-۲- جمع بندی؛ روش های طراحی سیستم های تصویرمحور	۳۵
۳- مفاهیم و ابزارها	۳۶
۱-۳- مقدمه	۳۷
۲-۳- ویژگی های محلی	۳۸
۱-۲-۳- ویژگی های محلی چه هستند؟	۳۸

۳۸ چرا ویژگی های محلی
۳۹ سیستم های شناسایی مبتنی بر ویژگی های محلی
۳۹ آشنایی با چند اصطلاح
۳۹ ۱-۴-۳ آشکارساز
۳۹ ۲-۴-۳ تغییرناپذیر یا همگرد
۴۰ ۳-۴-۳ تغییر ناپذیر با دوران یا ایزوتروپیک
۴۰ ۴-۴-۳ نقطه کلیدی ناحیه و ویژگی محلی
۴۱ ۵-۴-۳ خصوصیات ویژگی محلی ایده آل
۴۳ ۶-۴-۳ آشکارساز گوشه
۴۷ ۷-۴-۳ آشکارسازهای حباب
۵۰ ۸-۴-۳ تبدیل افاین
۵۲ ۵-۳ بررسی روش SIFT
۵۲ ۱-۵-۳ مقدمه ای بر روش SIFT
۵۵ ۲-۵-۳ آشکارسازی اکستریم های فضا-مقیاس
۶۲ ۳-۵-۳ تعیین محل دقیق نقطه کلیدی
۶۶ ۴-۵-۳ تخصیص جهت
۶۸ ۵-۵-۳ توصیف کننده محلی تصویر
۷۱ ۶-۵-۳ کاربرد در شناسایی اشیاء
۷۲ ۷-۵-۳ انطباق نقطه کلیدی
۷۴ ۶-۳ جمع بندی
۷۵ روش پیشنهادی
۷۶ ۱-۴ مقدمه
۷۶ ۱-۱-۴ اجزای اصلی سیستم تشخیص زبان ایما و اشاره
۷۶ ۲-۴ نحوه استفاده از SIFT در این تحقیق
۷۶ ۱-۲-۴ برنامه کاربردی siftWin32.exe
۷۸ ۲-۲-۴ نرم افزار PSLR
۷۹ ۳-۲-۴ ساختار کلی سیستم
۸۰ ۴-۲-۴ آماده سازی پایگاه داده
۸۱ ۵-۲-۴ آنالیز تصاویر آزمایشی

۳-۴- جمع بندی	۸۳
۵- آزمایشات، شبیه سازی و نتایج	۸۴
۱-۵- مقدمه	۸۵
۲-۵- مجموعه داده	۸۵
۱-۲-۵- زبان ایما و اشاره فارسی	۸۵
۲-۲-۵- روش جمع آوری تصاویر	۸۶
۳-۵- آزمایشات و نتایج آن ها	۸۷
۱-۳-۵- تأثیر نحوه تبدیل RGB به خاکستری، روی تعداد نقاط کلیدی استخراج شده	۸۷
۲-۳-۵- تأثیر پیش پردازش های تصاویر بر روی تعداد نقاط کلیدی استخراج شده از هر تصویر	۸۹
۳-۳-۵- تأثیر مقدار فاصله ی نخستین نزدیکترین همسایگی و دومین نزدیکترین همسایگی بر روی میزان تشخیص در تصاویر عادی و نویزی شده	۹۰
۴-۳-۵- تأثیر استفاده از فیلترهای حذف نویز بر روی دقت تشخیص در تصاویر نویزی شده	۹۲
۵-۳-۵- نمونه هایی از انطباق های صحیح و اشتباه	۹۳
۴-۵- جمع بندی نتایج	۹۵
۶- نتیجه گیری	۹۶
۱-۶- مقدمه	۹۷
۲-۶- مقایسه سیستم ارائه شده با تحقیقات انجام شده و نتیجه گیری	۹۷
۶-۳- کارهای آینده	۹۸
مراجع	۹۹

فهرست جداول

- جدول ۱-۲ : انتخاب روش شناسایی لبه‌ی مناسب ۲۱
- جدول ۲-۲ : تأثیر آستانه‌گیری روی برخی حروف ۲۲
- جدول ۳-۲ : ساختار اطلاعات نهایی ماتریس توزیع شعاع θ ۲۴
- جدول ۱-۵ : تعداد نقاط کلیدی بدست آمده برای هر تصویر آموزش با استفاده از تبدیل (۱) ۸۷
- جدول ۲-۵ : تعداد نقاط کلیدی بدست آمده برای هر تصویر آموزش با استفاده از تبدیل (۲) ۸۸
- جدول ۳-۵ : تعداد نقاط کلیدی بدست آمده برای هر تصویر آموزش Crop شده ۸۹
- جدول ۴-۵ : میزان دقت سیستم به ازای مقادیر مختلف disRatio برای تصاویر آزمایشی عادی ۹۰
- جدول ۵-۵ : میزان دقت سیستم به ازای $disRatio = 0.7$ به تفکیک علائم ۹۱
- جدول ۶-۵ : میزان دقت سیستم به ازای مقادیر مختلف disRatio برای تصاویر آزمایشی نویزی شده ۹۱
- جدول ۷-۵ : میزان دقت سیستم به ازای مقدار $disRatio = 0.7$ با اعمال فیلتر میانه ۹۲
- جدول ۱-۶ : مقایسه سیستم ارائه‌شده و سایر سیستم‌ها ۹۷

فهرست اشکال

- شکل ۱-۲: استفاده از دستان برهنه برای شناسایی حرکات دست ۶
- شکل ۲-۲: ساختار سیستم تشخیص ۱۳
- شکل ۳-۲: ویژگی‌های ژست: (a) تصویر اصلی (b) مرکز سطح و جهت ژست ۱۴
- شکل ۴-۲: دنبال کردن مرز: (a) تصویر ژست (b) مرز مربوطه (c) مرز اصلاح شده بوسیله فیلتر (۴) ۱۵
- شکل ۵-۲: ویژگی‌ها: (a) بردارهایی که از مرکز سطح به یک نقطه روی مرز کشیده می‌شوند (b) ویژگی‌ها (طول بردارها) که به صورت نمودار میله‌ای نمایش داده می‌شوند. ۱۷
- شکل ۶-۲: شمای کلی سیستم ۱۹
- شکل ۷-۲: تبدیل تصویر رنگی به تصویر باینری ۲۰
- شکل ۸-۲: ساده‌سازی علامت "o" ۲۳
- شکل ۹-۲: مختصات در فضای تصویر ۲۴
- شکل ۱۰-۲: شبکه دسته‌بندی ۲۵
- شکل ۱۱-۲: بخشی از ماتریس هدف برای مجموعه آموزش ۲۶
- شکل ۱۲-۲: علائم ایما و اشاره آمریکایی مورد استفاده در سیستم ۲۸
- شکل ۱۳-۲: ساختار کلی سیستم ۲۹
- شکل ۱۴-۲: تبدیلات مورد استفاده برای آماده‌سازی تصاویر نهایی ۳۰
- شکل ۱۵-۲: تبدیل موجی گسسته دو بُعدی ۳۱
- شکل ۱۶-۲: نمودار درختی تجزیه موجی ۳۳
- شکل ۱۷-۲: زیر تصاویر بازسازی شده برای ژست "ظ" ۳۳
- شکل ۱۸-۲: شبکه عصبی MLP مورد استفاده برای مرحله دسته‌بندی ۳۴
- شکل ۱-۳: ویژگی‌های محلی ۳۸
- شکل ۲-۳: نحوه انتخاب ویژگی‌های محلی ۴۴
- شکل ۳-۳: نحوه انتخاب ویژگی‌های محلی ۴۴
- شکل ۴-۳: استفاده از ماتریس هسین برای استخراج ویژگی‌های محلی ۴۷
- شکل ۵-۳: استفاده از دترمینان ماتریس هسین برای استخراج ویژگی‌های محلی ۴۹
- شکل ۶-۳: ساختارهای حسابی ۵۰
- شکل ۷-۳: نحوه ساخت هرم فضا-مقیاس در آشکارساز SIFT ۵۷
- شکل ۸-۳: نحوه تعیین اکسترمم‌های محلی در هرم فضا-مقیاس ۵۸
- شکل ۹-۳: نمودارهای تعیین فرکانس نمونه برداری در مقیاس ۶۰
- شکل ۱۰-۳: نمودار تعیین فرکانس نمونه برداری در مکان ۶۱
- شکل ۱۱-۳: نقاط کلیدی استخراج شده با در نظر گرفتن آستانه‌گیری ۶۴

- شکل ۳-۱۲ : نمودار پایداری نقاط کلیدی با افزایش نويز
- شکل ۳-۱۳ : رو به طرف بالا کردن ناحیه اطراف نقطه کلیدی همراه با جهتش
- شکل ۳-۱۴ : تعیین هیستوگرام‌های جهت برای توصیف‌کننده سیفت
- شکل ۳-۱۵ : انطباق یک نقطه با نقاط موجود در پایگاه داده
- شکل ۳-۱۶ : نمودار تابع چگالی احتمال انطباق‌های درست و اشتباه
- شکل ۴-۱ : زیرمنوی Train در نرم‌افزار PSLR
- شکل ۴-۲ : زیرمنوی Simulation در نرم‌افزار PSLR
- شکل ۴-۳ : زیرمنوی Setting در نرم‌افزار PSLR
- شکل ۴-۴ : ساختار کلی سیستم
- شکل ۴-۵ : نمای کلی سلول اطلاعاتی تصاویر آموزشی و روند تشکیل آن
- شکل ۴-۶ : نمای کلی سلول اطلاعاتی تصاویر آزمایشی و روند تشکیل آن
- شکل ۴-۷ : (a) حرف "م" ، (b) حرف "ک"
- شکل ۵-۱ : الفبای ایما و اشاره فارسی
- شکل ۵-۲ : نمایش نقاط کلیدی بدست آمده برای علامت "ا"
- شکل ۵-۳ : تطبیق صحیح برای حرف "پ"
- شکل ۵-۴ : تطبیق صحیح برای حرف "ا"
- شکل ۵-۵ : تطبیق صحیح برای حرف "ذ"
- شکل ۵-۶ : تطبیق اشتباه حرف "ک" و "ج"
- شکل ۵-۷ : تطبیق اشتباه حرف "ه" و "ژ"
- شکل ۵-۸ : تطبیق اشتباه حرف "آ" و "أ"

فصل اول

مقدمه

تأثیر کامپیوتر بر زندگی انسان روزبه‌روز مهم‌تر و چشم‌گیرتر می‌گردد. در این باب تعامل انسان و کامپیوتر که در حال حاضر به طور فزاینده رو به افزایش است، از اهمیت بالایی برخوردار می‌باشد. با آنکه تعامل بین انسان و کامپیوتر در دهه‌های اخیر دارای پیشرفت‌های بسیاری بوده است، لیکن بسیاری هنوز ارتباط با کامپیوتر را دشوار می‌دانند. لذا طراحان در تلاشند تا این تعاملات را سریع‌تر، آسان‌تر و طبیعی‌تر سازند. برای حصول این امر تکنیک‌های تعامل انسان با انسان در حال وارد شدن به حیطه تعامل انسان با کامپیوتر می‌باشند.

در این میان حرکات دست^۱ یکی از رایج‌ترین و طبیعی‌ترین راه‌های تعامل انسان‌ها با یکدیگر است، که توجه بسیاری را به خود جلب کرده است. اما اهمیت استفاده از حرکات دست برای ارتباط زمانی روشن‌تر می‌گردد که زبان ایما و اشاره^۲ در نظر گرفته شود. زبان ایما و اشاره روش ارتباطی پایه‌ای میان افرادی می‌باشد، که از نواقص شنوایی رنج می‌برند. این زبان مجموعه‌ای از حرکات^۳، حالات^۴، اداها و حرکات صورت می‌باشد، که مطابق با حروف و کلمات در زبان‌های عادی برای ایجاد ارتباط یک فرد عادی با یک فرد ناشنوا مورد استفاده قرار می‌گیرد.

زبان ایما و اشاره فارسی، زبان کاملی است که از علائمی که توسط حالات و حرکات دستان و دیگر اجزای بدن مثل صورت و چشمان اجرا می‌گردند، استفاده می‌کند. این زبان تقریباً شمال ۱۰۷۵ حرکت و کلمه رایج است، که شامل حرکات استاتیک و دینامیک می‌باشد [۱ و ۲].

تشخیص زبان ایما و اشاره فارسی بر نوشتاری کردن آن متمرکز است، تا ارتباط میان افراد ناشنوا با جامعه‌ی شنوا را راحت‌تر سازد. برای برگرداندن زبان ایما و اشاره به زبان عادی و بالعکس به یک مترجم و مفسر نیاز است.

در این راه وجود چالش‌هایی چون تغییر نقطه دید^۵، تغییر شدت روشنایی^۶، هم‌پوشانی^۷، تغییر اندازه^۸، شلوغی و درهم برهمی زمینه^۹، امر استخراج ویژگی‌ها و به تبع آن شناسایی را دچار اختلال

¹ Hand movement

² Sign language

³ Gesture

⁴ Posture

⁵ View point variation

⁶ Illumination

⁷ Overlapping

⁸ Scale variation

⁹ Background clutter

می‌کنند. با توجه به چالش‌های ذکر شده، سیستمی می‌تواند در امر شناسایی حالات دست موفق باشد، که در مرحله استخراج ویژگی‌هایش، ویژگی‌های تغییرناپذیر متمایزی^۱ را از تصویر استخراج کند.

در فصل دوم این تحقیق، مطالعات انجام‌شده در حوزه سیستم‌های مترجم زبان ایما و اشاره، مورد بحث و بررسی قرار می‌گیرد. فصل سوم مفاهیم و ابزارهای مورد نیاز برای پیاده‌سازی سیستم را معرفی می‌کند. در فصل چهارم روش پیشنهادی برای تشخیص زبان ایما و اشاره فارسی معرفی می‌گردد. فصل پنجم به آزمایشات انجام‌شده و نتایج آن‌ها اختصاص دارد و در نهایت در فصل ششم نتیجه‌گیری ارائه خواهد شد.

^۱ Discriminate invariant feature (DIF)

فصل دوم

بررسی مطالعات انجام شده

در این فصل، به بررسی روش‌های ارائه شده توسط محققان برای تشخیص حالات و حرکات دست پرداخته خواهد شد و مزایا و معایب هر روش، مورد بررسی قرار خواهد گرفت.

تلاش‌هایی که برای تشخیص خودکار زبان ایما و اشاره صورت گرفته، از دهه‌ی ۹۰ قرن بیستم در این حیطه پدیدار شده‌اند. بسیاری از محققان در تلاش بوده‌اند تا حرکات دست را به حیطه تعاملات کامپیوتر و انسان وارد سازند. آن‌ها در تلاش‌اند تا از حرکات دست در برنامه‌های کاربردی مثل بازی-های کامپیوتری، تعاملات گرافیکی کامپیوتری، هندترکینگ^۱، کنترل تلویزیون، و ابزارهای اشاره‌ای بعنوان جایگزینی برای موس معمولی استفاده کنند [۳].

به طور کلی شناسایی حرکات دست در تعاملات کامپیوتری به دو روش اصلی روش‌های پوشش-محور^۲ و روش‌های تصویرمحور^۳ امکان‌پذیر است..

روش‌های پوشش‌محور، متکی به ابزارهای الکترومکانیکی می‌باشد، در حالی که در روش دوم با استفاده از تکنیک‌های پردازش تصویر، حرکات دست تشخیص داده می‌شوند. این روش که مشکل عدم راحتی روش اول را مرتفع و از طرفی تعاملات را طبیعی‌تر می‌سازد، روش تصویرمحور نامیده می‌شود. گونه‌ای از سیستم‌های تشخیص حرکت تصویرمحور متکی به استفاده از پوشش‌هایی که به شکل خاصی طراحی شده‌اند و با علامت‌گذارهای تصویری، به تشخیص حرکت دست کمک می‌کنند [۴]. استفاده از پوشش و مارکرها به طبیعی کردن هرچه بیشتر چنین سیستم‌هایی کمکی نمی‌کند. علاوه بر آن اگر از دستکش‌های رنگی استفاده شود پیچیدگی پردازش بالا می‌رود.

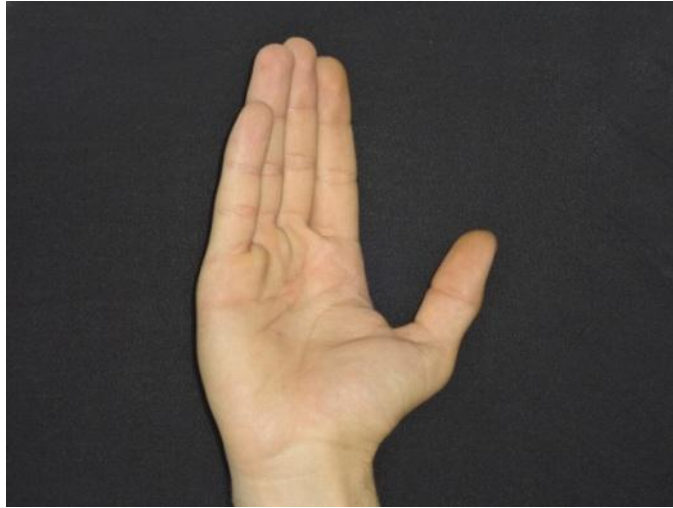
بنابراین سیستم‌های تشخیص حرکت تصویر محور تلاش می‌کنند تا راحتی و طبیعی بودن مورد نظر را با استفاده از تصاویر دستان^۴ برهنه برای تشخیص حرکات بدست آورد.

¹ Hand tracking

² Glove based

³ Image based

⁴ Bare hand



شکل ۱-۲: استفاده از دستان برهنه برای شناسایی حرکات دست

در ادامه ابتدا به بررسی کلی مطالعات انجام شده در دو حوزه پوشش محور و تصویر محور، به طور اجمالی پرداخته می شود و سپس چند روش رایج تر را به صورت مفصل بررسی می گردد. از آنجا که تمرکز این تحقیق بر روی روش های تصویر محور و دستان برهنه می باشد، لذا روش های مبتنی بر ابزارهای الکترومکانیکی و روش هایی که از دستکش های رنگی استفاده می کنند در حوزه روش های پوشش محور بیان می شوند.

۱-۲- تحقیقات انجام شده در حوزه روش های پوشش محور

۱-۱-۲- روش های مبتنی بر ابزارهای الکترومکانیکی

در این دسته که متکی به استفاده از ابزارهای الکترومکانیکی است، برای اندازه گیری پارامترهای حرکتی مختلف مثل موقعیت دست، زاویه و مکان سرانگشتان، از اطلاعات سنسورها استفاده می گردد. مشکلات عمده ی چنین سیستم هایی این است که کاربر را مجبور به پوشیدن ابزارهای دردسرساز و غیرراحت می کنند. این امر سبب پیچیدگی و غیر طبیعی بودن سیستم می گردد.

در سال ۱۹۹۶ و مپلو^۱ از یک Cyber glove با ترکیب مکانی به منظور تشخیص دقیق تر استفاده کرد. در همین سال تلاش های دیگری نیز در این حوزه انجام شد. رونگ هویی و او هیونگ^۲ از دستکش

^۱ Vamplew

^۲ Rung-Huei & Ouhyoung

Data glove به عنوان ابزار ورودی سیستم‌شان استفاده کردند. کادوس^۱ سیستمی که بر مبنای دستکش‌های برقی Power glove بود، طراحی کرد تا مجموعه ای از ۹۵ علامت مجزای زبان ایما و اشاره استرالیایی را تشخیص دهد. سیستم او قادر به تشخیص با دقت ۸۰٪ بود. مزیت روش کادوس کم‌بودن هزینه‌های محاسباتی آن بود [۵].

وگler و متاگراس^۲ در سال ۱۹۹۷ از دستکش‌های الکترومکانیکی و ترکرهای مغناطیسی برای گرفتن اطلاعات ورودی استفاده کردند. آنها در سیستم خود از روش‌های تصویر کامپیوتری برای انتخاب پارامترهای^۳ بُعدی حرکات بازوی کاربر و مدل‌های مخفی مارکوف^۴ برای تشخیص جملات زبان ایما و اشاره آمریکایی^۵ با مجموعه ۵۳ علامت استفاده کردند. آنها برای کاهش اثرات الحاقات حرکتی، مدل‌های مخفی مارکوف متن محور را طراحی کردند. دقت سیستم آنها ۸۹/۹٪ بود [۵]. در سال ۲۰۰۴ آنها سیستم خود را بهبود دادند [۶].

در سال ۱۹۹۷ چان سولی^۵ یک سیستم را برای تشخیص حرکات در زبان ایما و اشاره کره‌ای^۶ ارائه کرد. او از منطق فازی^۷ برای دسته‌بندی مسیر و از شبکه‌های عصبی^۸ min-max فازی برای تشخیص حرکت و جهت آن (جهت‌یابی) استفاده کرد. میزان متوسط تشخیص ۸۰/۱٪ بود. این سیستم، در زمره سیستم‌های تشخیص پوشش محور قرار می‌گیرد، چرا که از واحد میانجی Cyber GloveTM برای اندازه‌گیری خمش انگشتان، حرکت دست و جهت حرکت استفاده می‌کند [۳].

لیانگ و اوهیونگ^۹ از مدل‌های مخفی مارکوف برای تشخیص زبان ایما و اشاره تایوانی^{۱۰} استفاده نمودند. سیستم آنها ۲۵۰ علامت را شناسایی می‌کرد. آنها برای جمع‌آوری اطلاعات از Data glove به عنوان ابزار ورودی استفاده کردند. یکی از معایب سیستم آنها این بود که کاربر می‌بایست حرکات را آهسته انجام می‌داد. این تحقیق در سال ۱۹۹۸ انجام گرفت [۵].

¹ Kadous

² Vogler & Metaxas

³ Hidden Markov Models (HMMs)

⁴ American Sign Language (ASL)

⁵ C.-S. Lee

⁶ Korean Sign Language (KSL)

⁷ Fuzzy

⁸ Neural networks

⁹ Liang & Ouhyoung

¹⁰ Taiwan sign language (TSL)

در سال ۲۰۰۱ وانگ، جائو و زوان^۱ سیستمی بلادرنگ برای تشخیص جملات زبان ایما و اشاره^۲ چینی طراحی کردند. اطلاعات با دو Cyber glove و یک ترکر^۳ بُعدی جمع‌آوری می‌شد. آن‌ها مدل‌های مخفی مارکوف را با الگوریتم State tying، شناسایی مستمر فریم و الگوریتم جستجو به کار بردند [۷].

در سال ۲۰۰۳ چن، جائو، یانگ، فانگ و وانگ^۴ از دستکش‌های الکترومکانیکی و ترک‌های مغناطیسی برای جمع‌آوری اطلاعات استفاده نمودند. آن‌ها نیز از مدل‌های مخفی مارکوف برای تشخیص زبان ایما و اشاره چینی بهره گرفتند [۸].

۲-۱-۲- روش‌های مبتنی بر دستکش‌های رنگی

این دسته متکی به استفاده از دستکش‌های طراحی شده به همراه مارک‌های تصویری می‌باشد که اصطلاحاً حرکت تصویر محور با مارک‌های پوششی^۴ نامیده می‌شوند. این کار به مشخص کردن حالات دست کمک می‌کند.

گروبل و عصان^۵ در سال ۱۹۹۶ از مدل‌های مخفی مارکوف برای تشخیص استفاده کردند. سیستم آن‌ها ۲۶۲ علامت را با دقت ۹۱/۳٪ تشخیص می‌داد. برای جمع‌آوری تصاویر کاربران دستکش رنگی پوشیده بودند. در همان سال نام و وُهن^۶ از سیستم‌های تشخیص پوششی و ابزارهای علامت‌گذار تصویری برای تشخیص زمان و مکان حرکات دست استفاده نمودند. در این سیستم نیز از مدل‌های مخفی مارکوف بهره گرفته شده است [۵].

در سال ۱۹۹۸ یوشینوری، کانگ هیون، نوبوتاکا و یوشیاکی^۷ از دستکش‌های رنگی استفاده کرده و نشان دادند که استفاده از دستکش‌های رنگی صلب، امکان سریع‌تر شدن فرآیند انتخاب ویژگی نسبت به دستان برهنه را فراهم می‌آورد [۹].

¹ Wang, Gao & Xuan

² Chinese Sign Language (CSL)

³ Chen, Gao, Fang, Yang & Wang

⁴ Visual-based gesture with glove-markers (VBGwGM)

⁵ Grobel & Assan

⁶ Nam & Wohn

⁷ Yoshinori, Kang-Hyun, Nobutaka & Yoshiaki

حسین^۱ در سال ۱۹۹۹ سیستمی را برای تشخیص الفبای زبان ایما و اشاره عربی طراحی کرد که مبتنی بر تشخیص سر انگشتان و میچ دست با استفاده دستکشی رنگی شامل ۶ رنگ متفاوت بود. پس از تشکیل بردارهایی بین سر انگشتان و میچ اطلاعات بردارها وارد مجموعه‌ای از مدل‌های ANFIS^۲ می‌شد، تا امر تشخیص حرکت حاصل شود. عیب روشن این سیستم وابستگی‌اش به استفاده از دستکشی رنگی می‌باشد که محدودیتی را برای طبیعی بودن تعامل ایجاد می‌کند [۴].

۲-۲- تحقیقات انجام شده در حوزه روش‌های تصویر محور محض^۳

نوع دوم سیستم‌های تشخیص حرکت دست تصویر محور، مبتنی بر تصاویر دستان برهنه می‌باشد، که حرکت تصویر محور محض نامیده می‌شود. (حرکت تصویر محور بدون دستکش و مارکرها). این نوع برآن است که راحتی و طبیعی بودن مورد نظر را با استفاده از تصاویر دستان برهنه بدست آورد، تا بتواند حرکات را تشخیص دهد.

چارایافان و ماربل^۴ در سال ۱۹۹۲ روشی را با استفاده از پردازش تصویر برای درک زبان ایما و اشاره‌ی امریکایی مورد تحقیق و مطالعه قرار دادند. سیستم آن‌ها به درستی قادر به تشخیص ۲۷ علامت از میان ۳۱ علامت زبان ایما و اشاره آمریکایی بود. در سال ۱۹۹۵ استارنر و پنتلند^۵ از یک روش منظر-محور^۶ با یک دوربین برای انتخاب ویژگی‌های دو بُعدی به عنوان ورودی برای مدل‌های مخفی مارکوف، استفاده کردند. میزان صحت در تشخیص جملاتی که شامل ۴۰ علامت بودند، ۹۱٪ بود [۵].

فریمن و راث^۷ در سال ۱۹۹۵ از اطلاعات جهت‌یابی محلی که در تصویر حرکت جا داده شده بود، برای تشخیص حرکات دست استفاده کردند. آن‌ها از هیستوگرام‌های جهت‌یابی بعنوان یک بردار شاخص برای تشخیص حرکت استفاده کردند [۳].

یانگ و آهو جا^۱ در سال ۱۹۹۹ روشی را برای تشخیص حرکات دینامیک مورد تحقیق قرار دادند. به این صورت که آن‌ها از شناسایی رنگ پوست و تبدیلات Affine^۲ مناطق پوستی در حال حرکت برای

¹ Hussain

² Adaptive neuro- fuzzy inference system

³ Pure visual based gesture (PVBG)

⁴ Charayaphan & Marble

⁵ Starner & Pentland

⁶ View based

⁷ Freeman & Roth