

تشخیص کلام با استفاده از تکنیک‌های بینایی (کلمات محدود)

هادی سلطانی زاده^۱، نجمه زارع^۲

عضو هیات علمی دانشگاه سمنان، دانشکده برق و کامپیوتر

^۱ h.soltanizadeh@semnan.ac.ir

^۲ n.zare.98@gmail.com

چکیده

امروزه تشخیص کلام و صحبت جایگاه ویژه‌ای را در ارتباط انسان و ماشین ایفا می‌نماید. مقاله حاضر تکنیک جدید لب‌خوانی را که بر مبنای تصویر بنا شده را مورد استفاده قرار داده است. در این روش تصاویر متوالی از چهره و لبها دریافت شده سپس تصویر لب‌ها جدا گردیده و ویژگی‌های مربوطه استخراج می‌شود. با جمع بندی ویژگی‌ها و دسته بندی آنها با استفاده از شبکه‌های عصبی مختلف، کلام ادا شده مورد باز شناسی قرار خواهد گرفت. ویژگی‌های استخراج شده و تعیین موقعیت لبها از تصویر چهره از ویژگی‌های این تحقیق می‌باشد که سبب گردید تا درصد شناسایی در حدود ۹۲/۴ درصد برای فاز آموزش و ۸۶/۴ درصد برای فاز آزمایش بدست آید.

کلمات کلیدی

پردازش تصویری صحبت، شبکه‌های عصبی، بردار ویژگی‌ها

شبکه‌های عصبی ۶- ایجاد بانک‌های اطلاعاتی و دسته‌های مناسب ۷- ورود نمونه‌های جدید و تشخیص کلام.

۱- مقدمه

مطالعات بسیاری در زمینه پیدا نمودن روش لب‌خوانی مناسب، برای کلام به صورت هوشمندانه صورت گرفته است. Sumbly و Pollack در سال ۱۹۵۴، Neely در سال ۱۹۵۶، Erber در سال ۱۹۶۹، Binnie در سال ۱۹۷۴ و Erber در سال ۱۹۷۵ به مطالعه در این زمینه با بررسی چهره طبیعی و در سطوح نوین مختلف پرداخته‌اند [۱].

در سال ۱۹۸۵ فردی به نام McGrath و پس از آن فرد دیگری با نام Summerfield et. al در سال ۱۹۸۹ نشان داد که لبها بیش از نیمی از اطلاعات گفتاری که توسط چهره فرد گوینده ارسال می‌شوند را در بر دارند. در حالیکه بخش زیادی از این اطلاعات با رویت دندانها می‌تواند کامل شود. لذا در این تحقیق لبها به شکل مجزا مورد بررسی قرار خواهند گرفت.

مراحل کلی تشخیص تصویری کلام که در شکل (۱) آورده شده است عبارتند از:

۱- خواندن از فایل‌های AVI و جدا کردن فریم‌های آن از یکدیگر ۲- پردازش بر روی هر فریم از تصاویر برای پیدا کردن محل لب در تصاویر ۳- استخراج ویژگی‌هایی برای دنبال کردن حرکات لب در طول مدت بیان کلمه ۴- استخراج بردارهای ویژگی‌ها برای هر کلمه ۵- آموزش بردارهای ویژگی‌ها توسط

