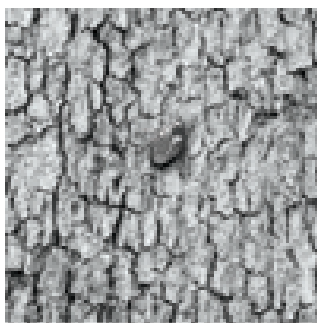


## پژوهش در زمینه بینایی کامپیوتری

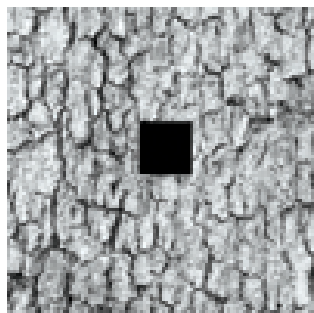
### مهرداد شهشهانی\*

دیگر معتبر نیست. یک مشکل عمده در استفاده از گشتاورهای مراتب بالاتر مجموعه‌های بزرگ داده‌ها، ناپایداری آنهاست. برای غلبه بر این مسائل، ما دو روش متمایز برای استخراج اطلاعات از یک تصویر عرضه کردیم که در آنها از جداسازی و تحلیل تغییرات ساختارهای همبستگی موضعی استفاده می‌شود. این روش‌ها، ناوردهایی عددی به دست می‌دهند که هم پایدارند و هم شامل اطلاعاتی درباره گشتاورهای بالاتر یک تصویر. کارایی این روش‌ها در مورد تشخیص یک شیء خارجی در یک تصویر، آزمایش شد. مثلاً شکل ۱، حشره کوچکی را روی یک درخت نشان می‌دهد و نمای (texture) حشره طوری است که با فضای اطراف کاملاً هماهنگ است.



شکل ۱

بدون استفاده از بانک داده‌ها یا حافظه و با تحلیل آماری موضعی تک تصویر، روش‌های ما مکان حشره را روی درخت، به طوری که در شکل ۲ دیده می‌شود به درستی، معین کرد. شاید نکته جالبتر این باشد که وقتی دو



شکل ۲

سکه به تصویر سکه‌ها (شکل ۳) اضافه کردیم، روش ما توانست سکه‌های جدید را تشخیص دهد (شکل ۴). به عبارت دیگر، این روش‌ها فقط نسبت



شکل ۴



شکل ۳

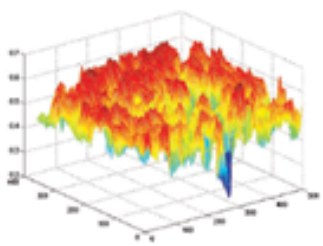
مسئله بینایی توجه مهندسان و دانشمندان از رشته‌های گوناگون را به خود جلب کرده است. به این مسئله می‌توان هم از دیدگاه عصب-روان‌شناختی و هم از دیدگاه محاسباتی نگریست: دیدگاه اول معطوف به فهم فرآیند بینایی در موجودات زنده است و دیدگاه دوم ناظر به ارتباط ادراک بصری انسان با کامپیوتر و در نتیجه، «بینایی کامپیوتری». عموماً پذیرفته‌اند که بیشتر فعالیت‌های مغز انسان بر فهم تصاویر بصری تمرکز دارد. ما کارهای ساده‌ای مانند تشخیص یک پروانه در میان انبوه گل‌ها یا حشره‌ای روی یک درخت یا به هم ربط دادن تصاویر مختلفی از یک شخص که از زوایای متفاوت و در زمان‌های گوناگون گرفته شده یا تشخیص سگ از گربه را بدیهی و پیش‌پا افتاده می‌دانیم ولی حتی با یک بررسی سطحی از فرآیند بصری که به چنین ادراکات پیش‌پا افتاده‌ای می‌انجامد، معلوم می‌شود که چه دشواری‌های بزرگی در فهم عملکرد مغز و نحوه دستیابی آن به این هدف‌ها وجود دارد. با ظهور روش‌هایی که از علوم طبیعی و ریاضی گرفته شده، پیشرفتی در کشف ماهیت برهم‌کنش‌های عصب-روان‌شناختی انسان به دست آمده است. با این حال، هنوز تا مرحله شناخت قطعی فرآیند بصری فاصله زیادی داریم.

مقایسه بینایی با گفتار ممکن است پیچیدگی‌های فرآیند بصری را روشن سازد. گفتار اساساً یک پدیده یک بعدی است ولی بینایی دست کم دوبعدی و احتمالاً سه بعدی است. با مطالعه حروف مجاور هم در یک زبان طبیعی می‌توان برنامه‌هایی کامپیوتری طراحی کرد که قطعه‌های جدید موسیقی یا کلمه‌ها و جمله‌های جدید تولید کند. حاصل کار، جمله‌های واقعی در یک زبان طبیعی یا قابل مقایسه با بهترین قطعات موسیقی نیست، ولی شبیه آنهاست. در بینایی، مسئله بسیار مشکل‌تر است. تلاش‌هایی برای تقلید فرآیند دیدن انجام شده اما نتایج آنها به هیچ وجه رضایت‌بخش نبوده است.

هرچند این دوجنبه یا دیدگاه یعنی دیدگاه عصب-روان‌شناختی و دیدگاه محاسباتی، به هم مربوط‌اند ولی اساساً متعلق به شاخه‌های متفاوتی از علم هستند. از تحقیق در هر یک از دو جنبه مسئله بینایی، مسائل چالش برانگیز زیادی سربرآورده است و در آینده قابل پیش‌بینی، انتظار می‌رود این دو حوزه همچنان حوزه‌های پر جنب و جوشی در علم و مهندسی باشند. حتی مسائل ساده تشخیص و تمایز هنوز راه‌حل‌های قطعی ندارند. پژوهش در زمینه بینایی در پژوهشکده ریاضیات پژوهشگاه معطوف به بینایی کامپیوتری است نه مسائل عصب-روان‌شناختی.

تحلیل آماری و کلاً تحلیل ریاضی داده‌های تشکیل دهنده یک تصویر رویکردی کلی است که گروه بینایی کامپیوتری در پژوهشگاه در پیش گرفته است. زمانی بلا جولس (Bela Julesz) یکی از پیشگامان مبحث بینایی، حدس زد که ادراک مستقیم انسان از یک تصویر تماماً مبتنی بر گشتاورهای اول و دوم داده هاست. ولی این فرض از نظر متخصصان بینایی کامپیوتری

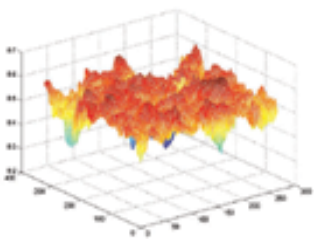
هم متمایز می‌کند. شکل‌های ۱۱ تا ۱۴ نشان دهنده تصاویری از یک سگ و یک گربه و رویه‌های مشخصه متناظر است.



شکل ۱۲



شکل ۱۱



شکل ۱۴



شکل ۱۳

این رویه‌ها در اساس ساختار نمای سگ‌ها و گربه‌ها را مشخص می‌کند. این البته با ادراک انسانی تفاوت بنیادی دارد زیرا ادراک انسان برای تمایز بین سگ و گربه به ساختار شکلی و حافظه اتکا می‌کند. همچنین توجه کنید که اعوجاج تصویر چهره انسان خواه بر اثر دوران یا تغییر نورپردازی موضعی یا وارد کردن عینک، تأثیری بر تعیین مکان چشم‌ها به صورتی که در شکل‌های ۱۵ و ۱۶ نشان داده شده، ندارد.



شکل ۱۶



شکل ۱۵

مسائلی که شرح آنها در بالا آمد فقط نمونه کوچکی از مسائلی هستند که نظر اهل علم و مهندسان را به سمت بینایی کامپیوتری جلب کرده‌اند. بسیاری از مسائل بنیادی در بینایی کامپیوتری به یافتن روش‌هایی برای سازماندهی داده‌های بصری فرو می‌کاهند به نحوی که مجموعه‌های داده‌های نزدیک به هم ارتباط نزدیکی با هم داشته باشند و اشیای متمایز نسبت به متریکی که به طور کارآمدی قابل محاسبه است دور از هم قرار گیرند. هرچند در این زمینه ایده‌های جالب توجهی مطرح شده، هنوز پاسخی قطعی به دست نیامده است.

\*\*\*\*\*

\* مهرداد شهشهانی، استاد پژوهشکده ریاضیات، پژوهشگاه دانش‌های بنیادی.

به بعضی از مشخصات ساختاری و کیفی تصویر حساس‌اند و نه به جزئیات ساختار. این نوع اطلاعات کیفی به راحتی قابل تشخیص با دید انسانی و قابل بیان در یک زبان طبیعی است ولی فرمولبندی همین پدیده در یک زبان کامپیوتری بسیار مشکل است.

یک مسأله مهم در بینایی کامپیوتری، جداسازی یا تفکیک یک شیء از محیط اطراف آن (segmentation) است. از روش‌های به دست آمده در پژوهشگاه می‌توان در مسأله جداسازی استفاده کرد؛ چند مثال را در شکل‌های ۵ تا ۱۰ می‌بینید.



شکل ۶



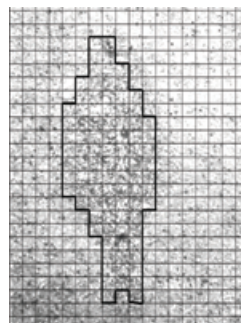
شکل ۵



شکل ۸



شکل ۷



شکل ۱۰



شکل ۹

توجه کنید که اسکی باز و وسائلس به درستی مشخص شده‌اند ولی سایه‌اش هرچند که از لحاظ رنگ متمایز از زمینه است، از زمینه تفکیک نشده است. دلیلش این است که عامل تعیین کننده اصلی، ساختار نمای تصویر است و نه رنگ‌ها یا مقدار پیکسل‌ها. همین‌طور، علی‌رغم فراوانی نوارهایی که رنگ‌های بسیار متفاوتی دارند، ببر به درستی تفکیک شده است. از موارد دیگری که ما روش‌های خود را در آنها آزمایش کردیم، تعیین مکان چشم‌ها در تصویری از چهره انسان و مسأله ساده ولی جالب بازگویی تفاوت بین یک سگ و یک گربه است. در مسأله دوم، رویه‌های مشخصه خاصی را به تصاویر سگ‌ها و گربه‌ها منسوب ساختیم و از طریق تحلیل نوسان‌های این رویه‌ها یک معیار عددی عملی به دست آوردیم که آنها را از